ORSAY

n° d'ordre

249

# SUR CERTAINS ASPECTS FONCTIONNELS DE L'AUDITION

---

Article Principal

recouvrant en partie la thèse de Doctorat d'Etat enregistrée au C.N.R.S. sous le numéro AO 1025

de

Jean-Claude RISSET

soutenue le 29 Mai 1967 à la Faculté des Sciences d'Orsay, devant le Jury :

MM. P. GRIVET
E. ROUBINE
A. BLAQUIERE
BENZECRI

A. JOLIVET )
MOLES ) invités

Extrait des ANNALES DES TÉLÉCOMMUNICATIONS
Tome 23. N° 3-4 — Mars Avril 1968.

# SUR CERTAINS ASPECTS FONCTIONNELS DE L'AUDITION *

par

Jean-Claude RISSET

Attaché de recherche au C.N.R.S. **

SOMMAIRE. — *Cet exposé comporte quatre parties : la première traite de modèles de la fonction auditive, la seconde de l'analyse en fréquence et de la perception des hauteurs sonores, la troisième des mécanismes de corrélation et de leur intervention dans l'audition, et la quatrième de la reconnaissance des formes telle qu'elle peut intervenir dans les systèmes artificiels ou dans la perception auditive.*

PLAN. — • *Introduction.* • 1. ***Fontion auditive et modèles*** — 1.1. *Quelques modèles partiels ;* — 1.2. *Modèle de l'audition proposé par A. Moles ;* — 1.3. *Quelques propriétés dont le modèle peut rendre compte ;* — 1.4. *Situation de l'audition humaine par rapport aux modèles.* • 2. ***Analyse en fréquence et perception des hauteurs sonores*** — 2.1. *Paradoxe de l'analyse en fréquence ;* — 2.2. *Affinement dans la perception des hauteurs ;* — 2.3. *Trois expériences sur la perception des hauteurs ;* — 2.4. *Remarques sur la perception des hauteurs musicales.* • 3. ***Corrélation, autocorrélation et audition*** — 3.1. *Cas d'un phénomène stationnaire ;* — 3.2. *Cas d'un phénomène variable au cours du temps ;* — 3.3. *Une méthode d'analyse en fréquence réalisable sur calculateur ;* — 3.4. *Autocorrélation et corrélation dans l'audition.* • 4. ***Reconnaissance des formes*** — 4.1. *Cas des systèmes artificiels : modèles pour machines ;* — 4.2. *Perception auditive et reconnaissance des formes.* • *Conclusion.* • *Bibliographie* (177 *réf.*).

## INTRODUCTION

Cet article résulte d'un travail surtout documentaire et théorique sur certains aspects fonctionnels de l'audition [1]. Nous y abordons des questions très diverses : plusieurs ne sont qu'effleurées, mais le texte renvoie à des références qui en traitent de façon détaillée, ce qui aidera peut-être à donner de l'audition une idée qui ne soit pas trop partielle.

Toutes ces questions, de l'analyse en fréquence à la psychophysique, de l'acoustique musicale à la reconnaissance des formes, gravitent autour d'un modèle de la fonction auditive proposé par A. Moles en vue d'une simulation sur calculateur électronique : elles se rapportent surtout à des aspects fonctionnels de l'audition, pour lesquels il est insuffisant de considérer l'auditeur comme un récepteur physique brut pourvu de « courbes de réponse » adéquates.

Bien que notre étude soit surtout théorique, nous avons utilisé un calculateur électronique pour illustrer un modèle simple d'affinement (cf. § 2.2) et pour engendrer des sons de caractéristiques physiques particulières (cf. § 2.3*a* et 2.3*b*). Les calculateurs électroniques peuvent en effet rendre de grands services en acoustique ; ils se distinguent naturellement par leur grande vitesse de fonctionnement, mais aussi par la possibilité de les assujettir à un programme qui peut comporter des ordres conditionnels et des boucles répétitives. Un calculateur peut donc effectuer des calculs, mais aussi simuler des mécanismes, des processus de décision, ce qui le rend apte à éprouver un modèle fonctionnel : on peut, utilisant un calculateur, expérimenter sur le modèle en modifiant les mécanismes simulés, bien plus facilement que si l'on effectue une simulation analogique faisant appel à des circuits ou des appareils dont la fonction est fixée à l'avance. Les calculateurs peuvent donc être précieux pour simuler des mécanismes fonctionnels comme ceux qui interviennent dans l'audition. D'autre part une onde sonore peut être décrite par une suite discrète d'échantillons numériques : un calculateur peut traiter ces nombres et, pourvu d'un équipement périphérique adéquat, analyser ou synthétiser des sons en permettant un contrôle rigoureux des paramètres. Nous avons tiré parti de ce nouveau moyen d'analyse et synthèse sonores dans une étude d'acoustique musicale décrite par ailleurs [3, 4, 5], ce qui nous a convaincu de son utilité et de sa puissance.

## 1. FONCTION AUDITIVE ET MODÈLES.

### 1.1. Quelques modèles partiels.

Ce que nous savons sur l'audition provient d'investigations dans plusieurs disciplines : anatomie, physiologie, psychologie. Les données anatomiques résultent de l'examen de la structure des systèmes auditifs ; les données physiologiques, de la mesure de la réponse physique de l'oreille et du système

* Partie d'un travail de thèse de Doctorat ès sciences physiques soutenue à la Faculté des sciences d'Orsay le 29 mai 1967.

** Institut d'électronique fondamentale, Faculté des Sciences, Orsay.

nerveux à des stimuli ; les données psychologiques, de la mesure de la réponse de l'organisme vivant tout entier à des stimuli. Dans une certaine mesure, ces méthodes s'excluent ; il est utile de les rapprocher et de les combiner [6] : c'est l'intérêt majeur des modèles. Un modèle peut donner lieu à des traitements mathématiques raffinés, et permettre des simulations (réalisables avec des moyens analogiques ou des calculateurs) grâce auxquelles il est possible d'étudier le comportement du modèle dans des conditions qu'on ne saurait reproduire sur une oreille ou un auditeur humain.

Ainsi Békésy, avec sa « camera acustica », a concrétisé un modèle de fonctionnement de la membrane basilaire dans l'audition, ce qui s'est révélé très fructueux [7]. Partant des résultats de Békésy, Flanagan a mis au point un modèle analytique détaillé de la réponse de la membrane basilaire, à partir duquel il a réalisé une simulation analogique à l'aide de circuits électriques [8], puis une simulation numérique à l'aide d'un calculateur [9] : ces simulations sont précieuses pour interpréter certaines expériences psychologiques [10, 11, 12], car elles indiquent comment l'information sonore est présentée aux fibres nerveuses.

Les phénomènes d'inhibition nerveuse latérale décrite par divers auteurs [7, 13, 14, 15], ont été simulés sur calculateur par Khanna et Noback [16] : ces phénomènes expliquent qu'un son puisse donner lieu à une décharge nerveuse plus sélective que la réponse de la membrane basilaire qui lui a donné naissance. (Nous en reparlerons en 2.1).

A un autre niveau, Mathews a défini un modèle qui permet des interprétations numériques précises de résultats d'expériences de détection auditive de signaux dans le bruit : les prévisions du modèle sont données par un calculateur convenablement programmé [6].

Licklider a proposé un modèle de « l'analgésie auditive », c'est-à-dire de la suppression de la douleur sous l'influence d'une stimulation acoustique appropriée : ce modèle très simple fournit un cadre cohérent pour l'étude du phénomène et suggère des expériences à entreprendre [17].

On pourrait citer bien d'autres modèles partiels qui ont aidé efficacement à mieux connaître certains aspects de l'audition. Nous allons maintenant décrire un modèle édifié par A. Moles, et qui se propose d'englober les divers niveaux de la fonction auditive.

### 1.2. Modèle de l'audition proposé par A. Moles.

Les théories de la physiologie de l'audition y sont reprises pour construire un modèle fonctionnel et non structurel, c'est-à-dire un modèle qui permette de simuler les fonctions de l'audition sans s'attacher à préciser les détails anatomiques et physiologiques qui assurent ces fonctions chez un auditeur réel [2].

Les divers niveaux du fonctionnement auditif sont symbolisés sous la forme suivante. (Fig. 1).

#### *a) Transduction.*

Le signal acoustique est transformé en signal électrique.

#### *b) Analyse en fréquence.*

Le signal électrique précédent est analysé par une série de filtres passe-bande en parallèle dont l'ensemble couvre les fréquences audibles : la largeur de bande de chaque filtre correspond à la surtension des résonateurs imaginés par Helmholtz [18, 19], répartis dans l'oreille interne.

#### *c) Envoi dans le nerf acoustique d'impulsions quantifiées.*

On fait intervenir ici la loi du « tout ou rien » qui régit la conduction nerveuse : on admet qu'un processus détecte l'énergie dans chaque filtre et que des impulsions, qui transportent sous forme codée l'information correspondante, sont envoyées dans les fibres nerveuses.

#### *d) Répartition des impulsions nerveuses.*

La théorie des fibres nerveuses dite « de la volée » [20] prévoit que les fibres peuvent se grouper et se

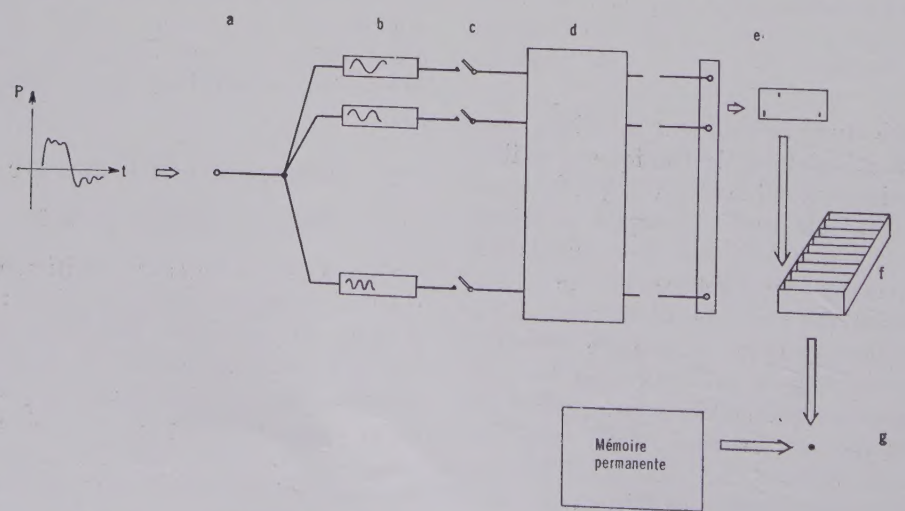


FIG. 1. — Schéma du modèle de l'audition proposé par A. Moles (explications dans le texte).



UNIV.

relayer à l'intérieur de chaque groupe pour transmettre des impulsions à un taux supérieur à celui qu'elles peuvent assurer individuellement (Fig. 2) : en effet, une impulsion parcourant une fibre l'inhibe pendant une période réfractaire de l'ordre de la milliseconde, tandis que des impulsions sont transmises au cerveau (jusqu'au complexe olivaire supérieur) à un rythme pouvant atteindre 3 000 ou 4 000 Hz [21]. Le modèle s'inspire de l'analogie de ce mécanisme à un multiplex télégraphique : un répartiteur permettra la transmission des impulsions à travers un certain nombre de voies.

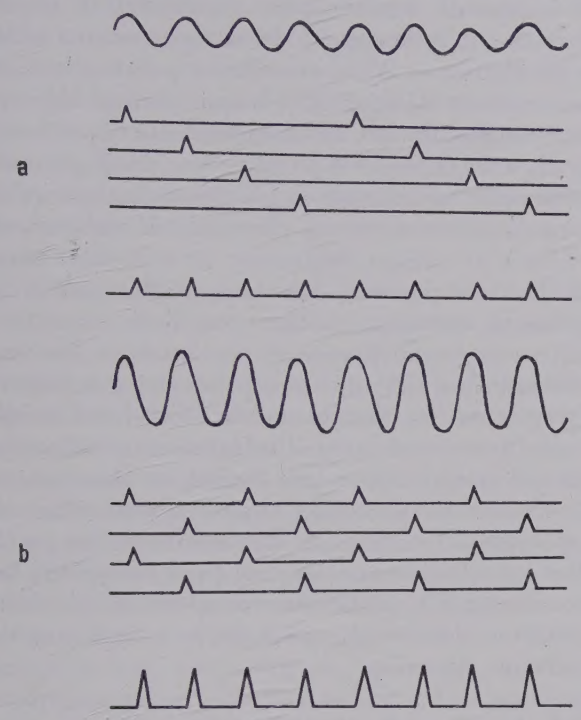


FIG. 2. — Schéma du mécanisme de la volée. En *a*) comme en *b*), on a figuré de haut en bas le signal sinusoïdal appliqué (de fréquence comprise entre 1 000 et 2 000 Hz), la réponse de chacune des quatre fibres qui se relaient pour transmettre les impulsions, et la somme de ces réponses, qui est synchrone au signal. *a*) et *b*) diffèrent par l'intensité du signal appliqué : quand cette intensité augmente, la fréquence de récurrence des impulsions sur chaque fibre croît (jusqu'à la limite imposée par la période réfractaire). Cette présentation est très simplifiée : ainsi l'intensité affecte également le nombre de fibres excitées.

***e) Perforation d'une carte tous les 1/16 de seconde.***

Il ne s'agit que d'une image mécanographique de l'intégration neurologique des données auditives : la carte perforée résume ce qui a été transmis par le nerf auditif durant l'intervalle de 1/16 s, sous forme d'un spectre instantané de puissance. A ce niveau, il n'y a plus de ressemblance manifeste entre le signal transmis et le signal d'entrée.

***f) Autocorrélation du phénomène.***

Les données auditives ne se résument pas à un flux de sensations successives : la cohérence du déroulement sonore est appréciée grâce à une mémoire à court terme, dite « mémoire phosphorescente » ou « durée de présence de la sensation » et qui s'étend sur quelques secondes. Cela se traduit dans le modèle par un mécanisme d'autocorrélation comparant la « carte » qui arrive aux cartes précédentes ; un phénomène sonore quasi stationnaire donnera un taux d'autocorrélation plus élevée qu'un phénomène transitoire. Cette autocorrélation correspond à des mécanismes décrits par Licklider [22, 23] ; elle porte sur un nombre limité de cartes correspondant à la capacité de la mémoire phosphorescente.

***g) Reconnaissance.***

Une dernière intégration conduit à la reconnaissance d'un « objet sonore » par corrélation de l'information transmise avec le contenu d'une mémoire à long terme ou mémoire « permanente ». Ce contenu est, chez l'auditeur humain, constitué au cours de l'expérience sensorielle et de l'éducation.

### 1.3. Quelques propriétés dont le modèle peut rendre compte.

Le modèle décrit au paragraphe 1.2 présente tout d'abord l'avantage de mettre en garde contre la tentation de considérer que les sons perçus aboutissent aux aires cérébrales avec les mêmes paramètres qu'on peut distinguer dans des ondes acoustiques [2] : les étapes schématisées (voir au § 1.2) imposent au signal des anamorphoses qui lui font perdre sa physionomie initiale.

D'autre part, le modèle comporte des mécanismes tels que *e*), *f*), *g*), nécessaires pour rendre compte des possibilités d'intégration de l'audition. On peut reconnaître une mélodie après transposition, identifier des instruments de musique d'après leur son, comprendre des phrases prononcées par différents locuteurs masculins ou féminins : l'audition humaine est apte à reconnaître des formes (§ 3), c'est-à-dire à classer des signaux auditifs complexes en catégories, chaque catégorie étant caractérisée non par des valeurs physiques brutes des signaux, mais par des relations entre leurs parties. De nombreuses données indiquent d'ailleurs qu'il existe plusieurs niveaux d'intégration auditive : il est d'expérience courante que d'une phrase entendue on retient à court terme les mots et à long terme le sens ; dans l'analyse musicale, on distingue différents niveaux d'organisation. Les expériences de Miller et Licklider [24] donnent des présomptions plus précises : étudiant l'intelligibilité $I(f)$ de la parole distordue en fonction de la fréquence $f$ d'interruption périodique, Miller et Licklider ont obtenu pour $I(f)$ deux minimums de l'ordre de 1 Hz et 300 Hz, respectivement (les valeurs exactes dépendent de la proportion de parole supprimée, entre 15 % et 75 % environ), qui pourraient correspondre à des valeurs critiques pour deux niveaux différents d'intégration perceptive.

Le modèle permet d'étudier les phénomènes de perception en présence de « bruit » dans la transmis-

sion nerveuse [25]. On peut y introduire des possibilités d'apprentissage auditif par modification du contenu de la mémoire permanente sous l'influence des signaux traités. Par exemple, on peut introduire *a priori* dans la mémoire quelques formes de courbes dynamique donnant le niveau en fonction du temps, correspondant à quelques timbres bien typés : la figure 3*a* correspond à une attaque « percutée », la figure 3*b*, à une attaque « éolienne » [26]. Si l'on fait

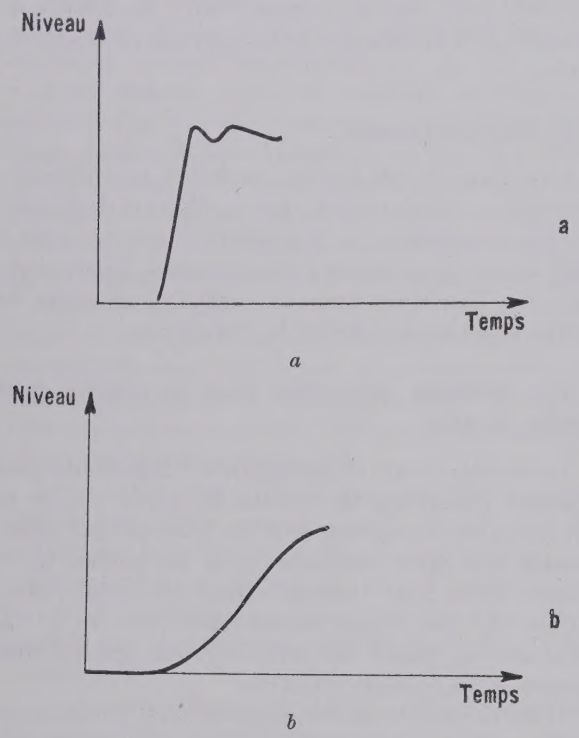


FIG. 3. — *a*) Courbe dynamique d'une attaque percutée (ou frappée) ; *b*) Courbe dynamique d'une attaque éolienne.

« écouter » à l'oreille simulée un son correspondant à une courbe nouvelle, la corrélation pourra identifier cette courbe à l'une des courbes connues, ou établir une relation entre cette courbe et les courbes connues. La répétition de ce son doit conduire à son introduction dans la mémoire comme nouvel élément caractérisé, comme nouvelle donnée de timbre. (Ainsi, le saxophone fut décrit à ses débuts comme mauvaise clarinette basse métissée de basson : on reconnut ensuite qu'il avait son timbre propre, digne d'enrichir la palette orchestrale). Le modèle permet donc de se faire une idée du rôle de l'apprentissage dans l'audition.

Les mécanismes d'intégration qui figurent dans le modèle permettent de rendre compte de phénomènes musicaux difficiles à expliquer par des considérations purement acoustiques [27]. On peut citer à cet égard le problème de l'appréciation des rapports harmoniques et mélodiques. La théorie de Helmholtz, qui attribue le sentiment de dissonance de deux sons simultané à des battements [18], se heurte à de nombreuses difficultés (la plus évidente tient à ce qu'elle attribue à un même intervalle harmonique des taux de dissonance hautement variables suivant les transpositions de cet intervalle ; la plus fondamentale tient à ce qu'elle attribue le sentiment de dissonance à des critères purement physiques et intrinsèques) : même si une théorie plus satisfaisante faisant intervenir la surtension de la membrane basilaire paraît devoir la remplacer [28], il reste à tenir compte de l'auditeur, de son conditionnement culturel [29] et de son état présent [30]. Il est connu en harmonie classique que la « préparation » de certaines notes d'un accord, c'est-à-dire l'introduction de ces notes antérieurement à l'accord complet, atténue le caractère dissonant éventuel de cet accord [31] ; pour interpréter les jugements harmoniques musicaux, il paraît nécessaire de faire appel à des notions comme celles de corrélation ou d'autocorrélation ; cette nécessité était apparue à Pikler [32]. L'appréciation des rapports mélodiques fait de façon évidente appel à une mémoire [33] ; Meyer a pu expliquer des jugements d'intervalles mélodiques en postulant une opération de comparaison nerveuse cherchant à réaliser une économie de codage [34].

Enfin il est possible, dans le cadre du modèle, de représenter certaines malfonctions de la perception auditive et particulièrement des agnosies (ou surdités centrales) [35] qui correspondent à des défauts d'intégration : en effet, le modèle décrit fonctionnellement des mécanismes d'intégration proches de ceux qui interviennent dans l'audition humaine.

Nous partons de ce modèle pour approfondir certains aspects fonctionnels de l'audition. En particulier les schémas analogiques qui symbolisent les niveaux auditifs constituent des références qui nous aideront à donner une idée de la complexité de l'audition humaine.

### 1.4. Situation de l'audition humaine par rapport aux modèles.

#### *a) Modèles complémentaires.*

Notons d'abord que les simulations fondées sur des modèles sont loin, le plus souvent, de reproduire la finesse de discrimination physique atteinte dans l'audition humaine, et se contentent d'une image grossière.

Mais la difficulté majeure des modèles réside dans l'apparente contradiction des données expérimentales sur l'audition, données qui se prêtent mal à entrer dans le cadre d'un modèle unique. Cela semble tenir à la complexité intrinsèque des mécanismes auditifs. Le système auditif humain paraît agencé d'une façon qui lui permet de donner des indications valables à partir de signaux imprécis, mais structurés (nous en reparlerons particulièrement en 4.2), alors que le détail de sa structure anatomique (et surtout de sa structure nerveuse) est imparfait ou variable d'un individu à un autre. Cette capacité de l'audition n'a rien de paradoxal à la lumière de résultats de la théorie de l'information : pour un canal de capacité $C$ en présence de bruit, et un signal de débit d'information $H$ inférieur à $C$, on sait qu'il existe un codage du signal permettant sa restitution

avec un taux d'erreur arbitrairement petit après transmission par le canal [36] ; et Moore et Shannon [37] ont montré qu'à partir de relais de faible fiabilité (décrite en terme de probabilités), on pouvait construire des circuits aussi fiables qu'on voulait, en agençant suffisamment de ces relais de façon adéquate. Van Bergeijk [38] a simulé l'innervation de la cochlée en couplant à une simulation électrique semblable à celle de Flanagan [8] des circuits « neuromimes » de Harmon [39], qui reproduisent certaines propriétés fonctionnelles des structures nerveuses (seuil, sommation, tout ou rien, réfractarité, inhibition) : le comportement de cette simulation est dans l'ensemble peu sensible au type exact de « neuromime » utilisé, ce qui montre qu'il est possible de construire des systèmes dont le fonctionnement n'est pas très sensible aux détails de structure. A. Moles proposait d'ailleurs d'illustrer, dans le cadre de son modèle [1], l'effet minime de la défaillance d'un élément sur le fonctionnement d'un système suffisamment redondant.

Il semble bien que, dans la perception auditive, plusieurs mécanismes différents puissent jouer, à la fois, ou en fonction des circonstances, pour tirer utilement parti d'indices distordus mais surabondants. Or un calculateur électronique fonctionne en série, traitant les informations les unes après les autres : cela ne permet guère de simuler fidèlement les opérations nerveuses et mentales d'êtres vivants, où un grand nombre d'informations sont traitées en parallèle pour exploiter au mieux leur redondance.

De même que certains modèles des noyaux atomiques, il est des théories ou des modèles auditifs, à première vue antinomiques, qui sont en réalité complémentaires : les uns ou les autres paraîtront mieux fondés suivant le type d'expériences qu'on cherche à interpréter. C'est le cas des théories de la hauteur dites théorie de la localisation et théorie de la périodicité (place pitch, periodicity pitch) (cf. § 2.1). Cette pluralité de modes de fonctionnement ne se laisse pas traduire aisément dans un modèle unique ; on peut songer à varier les paramètres de la simulation, mais cela amène à considérer un très grand nombre de cas. Faire un choix fixe de mécanismes ou de paramètres, c'est s'écarter de la manière dont la fonction auditive procède réellement : nous allons le montrer sur trois exemples précis.

### *b) Seuils différentiels.*

Si l'on se fonde sur la valeur des seuils différentiels en intensité pour déterminer une finesse de codage nerveux, l'on admet implicitement que ces seuils différentiels correspondent à des quanta perceptifs intrinsèques. Des expérimentateurs ingénieux ont pu, au cours d'expériences très soignées, mettre en évidence un seuil différentiel intrinsèque, correspondant à ce que Stevens appelle un « quantum neural » [7, 40, 41] : cela est très difficile, et l'interprétation des résultats a pu être mise en doute par certains auteurs [42, 43] ; il faut en effet, pour cela, des observateurs très attentifs et des conditions expérimentales très particulières permettant d'atteindre le plus directement possible les informations nerveuses accessibles à la conscience. Mais dans la plupart des circonstances expérimentales mises en jeu pour la mesure des seuils le « bruit » n'est pas suffisamment réduit, et les seuils mesurés n'ont qu'une signification statistique [44] : en effet, la réaction du sujet fait alors intervenir des mécanismes complexes, jouant à différents niveaux. Ainsi, dans une expérience de détection d'un signal très faible en présence de bruit, la performance d'un sujet sera moins bonne si l'on pénalise les erreurs (les détections par le sujet d'un signal alors qu'on ne lui présentait que du bruit) que si l'on valorise les détections exactes [45]. De même dans une expérience de mesure des seuils différentiels le sujet doit décider si les signaux qui lui sont présentés (éventuellement entachés d'un bruit spontané d'origine nerveuse [25]) sont différents ou non : le seuil différentiel qu'on mesure peut ne pas être intrinsèque, mais dépendre de la stratégie (hardie ou timorée) adoptée par le sujet pour prendre sa décision — cette stratégie dépendant elle-même des conditions de l'expérience. Nous décrirons au paragraphe 2.3*a* une expérience psychoacoustique indiquant que la faculté de discriminer entre deux signaux peut être influencée par le contexte, par les signaux présentés avant et après.

La coexistence de seuils intrinsèques et de seuils statistiques [46] (les derniers « noyant » les premiers) n'est pas pour faciliter la conception d'un modèle. Si l'on schématise le canal auditif par un canal technique où les informations seraient transportées sous forme d'impulsions codées, il faudrait en toute rigueur choisir, pour la séparation entre niveaux adjacents, une valeur correspondant aux seuils intrinsèques ; mais cela serait peu utile, car ces seuils intrinsèques sont le plus souvent masqués par la complexité de la réponse : les seuils rencontrés correspondent le plus souvent au résultat d'une opération effectuée à un niveau élevé et non à une quantification opérée dès l'entrée du canal auditif, ils ne traduisent pas une capacité ultime de détection, mais ils dépendent du processus adopté.

Ces considérations nous conduisent à ne pas admettre sans discernement les évaluations de la capacité d'information du canal auditif faites à partir des seuils différentiels. Certains auteurs [19, p. 153] ont divisé l'aire audible (dans le plan intensité × fréquence) en cellules censées représenter les « quanta acoustiques » que l'oreille peut distinguer : le nombre de ces « quanta acoustiques » est le produit du nombre d'échelons de fréquence discernables et du nombre d'échelons d'intensité discernables. Ces calculs de quanta acoustiques sont sujets à caution : il ne semble pas justifié de combiner des possibilités de discrimination élémentaire, car on s'écarte des procédures psychophysiques dont dépend la valeur des seuils. Mais surtout, il est absolument illégitime de calculer le débit d'information en mul-

tipliant le nombre de quanta acoustiques par le nombre d'échelons de durée discernables, calculé à partir d'une valeur admise pour la résolution temporelle de l'oreille : en effet discriminer deux sons de fréquences très voisines demande un certain temps, de l'ordre d'une fraction de seconde à plus d'une seconde — et plus la tâche est difficile, plus elle demande de temps ; en tout cas, on ne peut nullement admettre que l'oreille, discriminant des fréquences proches, conserve une résolution temporelle élevée : même la valeur de 0,25 s admise par Jacobson [47] est contestable. Il peut être utile de déterminer ou de mesurer des capacités informationnelles auditives dans certains cas précis [48, 49], mais les résultats ne s'étendent pas aisément à d'autres circonstances : par exemple des sujets doués d'oreille absolue peuvent, dans les conditions expérimentales définies par Pollack [48], traiter des débits d'information considérablement supérieurs aux autres sujets [50], alors que pour d'autres tâches auditives, on ne verra plus apparaître la différence entre sujets doués ou non d'oreille absolue. Comme nous l'a fait remarquer Benzécri, l'information au sens de Hartley et Shannon [36] et exprimée en bits est une notion mal adaptée à la communication sensorielle : l'additivité de l'information pour des événements indépendants est utile dans le domaine technique, mais pour la perception les événements sont rarement indépendants, ils sont liés hiérarchiquement et les informations qu'ils convoient peuvent s'activer mutuellement, ce qui rend caduque l'algèbre informationnelle.

### *c) Constante de temps.*

On conçoit souvent l'audition comme un processus physique ayant une certaine inertie et possédant une constante de temps. Il faut en effet un certain temps pour mettre en mouvement les parties mobiles de l'oreille : l'oreille moyenne agit comme un filtre passe-bas qu'on peut caractériser par une constante de temps mécanique. Harris [51] a pu mesurer une résolution temporelle auditive inférieure à 1 milliseconde, qui approche sans doute de près cette constante de temps mécanique. Cette valeur est beaucoup trop faible pour rendre compte des phénomènes d'intégration temporelle dans l'audition [52] : ce sont généralement des phénomènes nerveux et non plus des phénomènes mécaniques qui entrent en jeu, et Joos [53] avance la valeur de 50 millisecondes. Mais diverses observations indiquent qu'il est inadéquat de spécifier une constante de temps unique pour l'audition [52, 54]. Cette question est liée au pouvoir de discrimination en fréquence de l'oreille (cf. 2.1) : une batterie de filtres constitue un analyseur spectral à paramètres constants, qui est limité par $\Delta f . \Delta t \cong 1$, où $\Delta f$ et $\Delta t$ représentent les pouvoirs de résolution en fréquence et en temps ; si les filtres sont très sélectifs, ils ne transmettront pas les transitoires rapides, alors que l'oreille est capable de discriminer deux sons de fréquences très voisines aussi bien que deux transitoires très rapides. Gabor, analysant cette remarquable faculté, conclut que, suivant que la discrimination des signaux entendus se fait surtout en temps ou en fréquence, l'audition peut ajuster sa constante de temps entre 20 et 250 millisecondes au moins [54].

De nombreux exemples démontrent la complexité des caractèristiques temporelles de l'audition. Zwislocki [55] a proposé une théorie de la sommation auditive temporelle (c'est-à-dire de l'accroissement de la sensation quand on accroît la durée du stimulus) : il décrit ce phénomène en faisant intervenir une constante de temps de 200 millisecondes. Hirsch indique [56] que le délai minimal entre deux sons doit être de 2 millisecondes environ pour qu'un auditeur puisse détecter leur non-simultanéité, mais qu'il doit être de 20 millisecondes environ pour que l'auditeur puisse indiquer l'ordre de succession des deux sons. On peut rapprocher cette donnée de celle de Pikler [57] qui estime à 30 millisecondes le décalage maximal pouvant exister entre l'attaque de plusieurs instruments de musique pour que l'auditeur ait l'impression qu'ils sont ensemble. Des résultats que nous avons obtenus sur la reconnaissance de segments de sons de trompette [3, 5] montrent que des auditeurs peuvent effectuer des reconnaissances faisant appel à l'appréciation de détails temporels, même pour des signaux de durée réduite à 30 millisecondes. Enfin, étudiant la discriminabilité des temps de montée de deux sons purs de fréquence 1 000 Hz, Tenney [58] conclut que, pour que la différence soit perceptible, elle doit être au moins de l'ordre de 3 ms tant que les temps de montée ne dépassent pas 8 ms, et qu'elle doit correspondre à un rapport 3/2 au moins, pour des temps de montée supérieurs. Ces données, comme celles de Pollack [59], ne peuvent guère s'interpréter en admettant une constante de temps unique pour l'audition.

### *d) Capacité de la mémoire phosphorescente.*

Les expériences de Miller [60, 61] indiquent que la mémoire phosphorescente a une capacité assez bien définie et constante ($7 \pm 2$) si elle est exprimée non pas en bits, mais en objets (ou « items ») (chiffres, lettres, notes de musique...). Cela implique que la mémoire phosphorescente se constitue après un traitement préalable, une mise en forme élaborée de la perception, ce qui cadre mal avec l'hypothèse d'autocorrélations ou corrélations effectuées d'après le contenu de la mémoire phosphorescente.

Les conclusions de Miller, si elles sont commodes pour interpréter nombre d'expériences, sont toutefois insuffisantes pour rendre compte de la complexité de la mémoire phosphorescente : Crossman [62] considère qu'elle est intermédiaire entre une mémoire à capacité constante en bits et une mémoire à capacité constante en items. Récemment, on a pu mettre en évidence une chute brusque de la capacité de la mémoire auditive à court terme dans les premiers dixièmes de seconde [63] ; cet

effet correspond, pour l'audition, à celui qui a été révélé pour la vision par Sperling [64], et qui conduit à attribuer à la mémoire instantanée une capacité très supérieure aux estimations antérieures : mais une faible partie seulement de cette capacité peut être transférée dans une mémoire à court terme plus durable, et c'est cette partie qui correspond aux évaluations de Miller [60]. Enfin, Shepard, analysant la faculté qu'ont les sujets de reconnaître des stimuli visuels déjà présentés, en a déduit des estimations de la capacité de la mémoire humaine très supérieures à toutes les estimations antérieures [65]. Des conclusions semblables sont probablement valables dans l'audition ; elles impliquent que la mémoire phosphorescente puisse avoir un contenu très riche : les chiffres $7 \pm 2$ donnés pas Miller [60] pourraient alors ne traduire qu'une capacité de classement instantané des stimuli et non une capacité réelle de mémoire.

Nous allons à présent étudier le problème de la perception des hauteurs sonores, pierre d'achoppement des théories de l'audition, qui doivent expliquer comment l'oreille peut concilier un pouvoir élevé de discrimination en fréquence et un amortissement important.

## 2. ANALYSE EN FRÉQUENCE ET PERCEPTION DES HAUTEURS SONORES.

### 2.1. Paradoxe de l'analyse en fréquence.

La conversion des vibrations acoustiques en impulsions nerveuses se fait dans l'oreille interne au niveau de la membrane basilaire. Helmholtz [18] a noté l'analogie de cette membrane, formée d'environ 24 000 fibres transversales élastiques dont les dimensions diminuent régulièrement d'une extrémité à l'autre, avec l'ensemble des cordes d'un piano ou d'une harpe : il a attribué la faculté qu'a l'oreille de discriminer des sons de fréquences différentes et d'analyser des sons complexes en leurs partiels, à une analyse mécanique réalisée par l'ensemble de ces fibres, considérées comme des résonateurs accordés.

Cependant, la finesse de la discrimination des fréquences sonores que peut accomplir l'oreille pose un problème : elle suppose en effet que ces résonateurs mécaniques ont une forte surtension. Or, les fibres de la membrane basilaire, faite de tissus mous plongés dans des liquides visqueux, sont très amorties, à la différence des cordes d'un piano ou d'une harpe très peu amorties et qui continuent à vibrer après excitation. A la suite surtout des travaux de Békésy [7], on peut considérer que le comportement mécanique de la membrane basilaire est bien connu. On sait que les ondes sonores font vibrer le tympan ; le son est transmis par la chaîne des osselets à la base de la cochlée avec une efficacité de transmission qui décroît au-dessus de 3 000 Hz ; les muscles des osselets, par un processus réflexe, atténuent la vibration si elle est trop intense [66]. La membrane basilaire divise le limaçon en deux rampes : la vibration sonore fait naître des ondes dans la rampe supérieure ; ces ondes progressent jusqu'à l'apex et se réfléchissent dans la rampe inférieure. Il en résulte pour la membrane basilaire des déformations dont l'amplitude est maximale en un point dont la position dépend de la fréquence. Ce mécanisme d'analyse par la membrane basilaire a pu être confirmé par des enregistrements du potentiel microphonique cochléaire [67]. Il existe des simulations [8, 9] qui donnent de ce comportement une image très précise et qui permettent de considérer que la membrane basilaire a une caractéristique d'analyseur mécanique à large bande, avec une surtension inférieure à 2. De ce fait la réponse aux signaux transitoires est satisfaisante, mais l'analyse en fréquence de signaux stationnaires est très grossière, en contradiction apparente avec les facultés de l'audition.

De nombreuses théories ont été échafaudées pour tenter de lever cette contradiction : la première en date est sans doute la théorie téléphonique de Rutherford, publiée en 1886 [68], et suivant laquelle l'oreille, semblable à un téléphone, ne ferait que transmettre au cerveau une image électrique fidèle de l'onde sonore, le cerveau opérant toutes les analyses sans qu'il y ait localisation des fréquences au niveau de la membrane basilaire. Cependant cette hypothèse ne cadre pas avec les observations des influx du nerf auditif [21, 69, 70] et de nombreuses expériences vont à l'appui du principe d'une certaine localisation des fréquences : Békésy [7] a observé la position d'amplitude maximale suivant la fréquence sur de nombreux animaux ; Schuknecht et Sutton [71] ont provoqué des lésions sur la cochlée du chat et ont pu relier la position de ces lésions aux pertes auditives résultantes. Aussi sont apparues des théories de la localisation plus élaborées que celle d'Helmholtz : le principe d'une analyse mécanique au niveau de la membrane basilaire y est conservé, mais interviennent en outre des processus conduisant à une sélectivité accrue au niveau de l'excitation nerveuse. C'est ainsi que des connexions nerveuses d'inhibition peuvent entraîner un affinement de la distribution des décharges nerveuses par rapport à la distribution des amplitudes de vibration [7, 14, 15] : ce phénomène a pu être simulé sur un calculateur [16]. Huggins [72] a étudié de près l'interaction entre la membrane basilaire et la membrane tectoriale : c'est dans cette région que se trouvent les cellules ciliées, responsables de la transduction nerveuse qui code sous forme d'impulsions certaines caractéristiques du déplacement de la membrane basilaire. Huggins a montré que la densité du cisaillement provoqué par le déplacement de la membrane basilaire par rapport à la membrane tectoriale était proportionnelle à la dérivée quatrième de l'élongation par rapport à la position sur la membrane basilaire et que cela correspondait à une excitation beaucoup plus sélective que le déplacement de la membrane basilaire. Jacobson [19,

p. 153] propose une autre explication : l'oreille, comportant en parallèle des filtres peu sélectifs, peut néanmoins être sensible à de petites variations de fréquence dans la mesure où elles se traduisent par des variations d'amplitude dans les bandes correspondantes. Cette hypothèse d'un mécanisme commun pour la détection des variation de fréquence ou d'amplitude est corroborée par une expérience récente [73]. D'autres mécanismes encore ont été étudiés [74].

Toutefois, les théories de la localisation ne peuvent expliquer la perception de la hauteur des sons graves. (La hauteur sonore est un attribut psychologique étroitement lié à la fréquence fondamentale pour les sons périodiques, et qu'on définit comme permettant de ranger les sons sur une échelle grave-aigu ; nous reviendrons là-dessus au paragraphe 2.4). Un phénomène décisif à cet égard est celui de la perception des fondamentaux graves subjectifs : dans certaines conditions, un son complexe privé de sa composante fondamentale est perçu comme ayant la hauteur correspondant à la fréquence de cette composante fondamentale, c'est-à-dire la même hauteur que s'il n'était pas privé de cette composante. On a cru longtemps que l'audition des fondamentaux physiquement absents, bien connue des facteurs d'orgue, était due à la non-linéarité de l'oreille, reconstituant physiquement le fondamental à partir de ses harmoniques [75], de la même façon que certains appareils construits pour détecter la fréquence fondamentale de la parole [76]. Or la non-linéarité de l'oreille n'est sensible qu'aux fortes intensités [23], alors que le fondamental subjectif est perçu aux faibles intensités. Schouten a étudié systématiquement la perception des « fondamentaux résiduels » ou « résidus » qu'il définit ainsi : le résidu est la perception globale d'un certain nombre de composantes de Fourier voisines comme un son de timbre aigu et de hauteur grave [77]. Par exemple un son comportant des composantes sinusoïdales de fréquences 1 800 Hz, 2 000 Hz et 2 200 Hz est perçu comme ayant la même hauteur qu'un son sinusoïdal de fréquence 200 Hz. Cela n'est pas dû à la création physique, par non-linéarité de l'oreille, d'une vibration sinusoïdale de 200 Hz : en effet, on ne peut obtenir de battements avec des sons de fréquence voisine de 200 Hz ; d'autre part, un bruit capable de masquer le son sinusoïdal de fréquence 200 Hz ne masque pas le résidu, qui continue d'être entendu avec la même hauteur sonore. De plus, si l'on déplace de 40 Hz toutes les composantes le long de l'axe des fréquences (les composantes de fréquences 1 840 Hz, 2 040 Hz, 2 240 Hz ne sont alors plus harmoniques), le complexe est perçu comme ayant la même hauteur qu'un son sinusoïdal de fréquence 204 Hz [77]. Il ne s'agit donc pas d'un son différentiel ; il est clair, dans ce cas, que la hauteur sonore perçue ne correspond pas à l'espacement des raies spectrales, comme certains auteurs l'avaient supposé.

Pour expliquer le phénomène du résidu, il faut admettre que la détection de la hauteur fait appel à une analyse de type temporel pouvant apprécier la périodicité de l'onde sonore. La fonction :

$$f(t) = a_p . \cos(p\omega t + \varphi_p) + a_q \cos(q\omega t + \varphi_q),$$

est périodique et de période $2\pi/\omega$ si $p$ et $q$ sont premiers entre eux ; en particulier, si $p$ et $q$ sont des entiers consécutifs, la périodicité correspond à l'inverse de l'écartement des raies spectrales ; cela correspond à une fréquence $f_0 = \omega/2\pi$, bien que la composante fondamentale de fréquence $f_0$ soit physiquement absente. Le résidu résultant de composantes harmoniques s'explique si l'on admet que la périodicité du son détermine sa hauteur. Mais il reste à préciser le mécanisme de cette détection de la périodicité. Pour cela, il faut prendre en considération d'autres résultats expérimentaux sur la hauteur des sons pulsés : par exemple, une sinusoïde de 2 200 Hz pulsée à un rythme de 150 Hz donne lieu à la perception d'une hauteur grave correspondant à celle d'une sinusoïde de fréquence 150 Hz, bien que l'analyse de Fourier du son pulsé montre qu'il n'y a pas ou très peu d'énergie à 150 Hz [78]. Tonndorf [74] considère que la cochlée effectue une véritable démodulation du signal, détectant l'enveloppe ; son point de vue est une extension de la théorie de la localisation : la sinusoïde de fréquence 2 200 Hz, pulsée à la fréquence 150 Hz, induirait une activité nerveuse dans des canaux correspondant à la fréquence 150 Hz, ce qui expliquerait la perception d'une hauteur grave correspondant à celle d'une sinusoïde de fréquence 150 Hz. Or, comme dans le cas du résidu, on ne peut masquer cette hauteur grave par des sons capables de masquer une sinusoïde de 150 Hz [78], ce qui indique que l'activité nerveuse est en fait induite dans des canaux correspondant aux hautes fréquences : par conséquent, dans ce cas, la hauteur perçue n'est pas liée à la localisation de l'activité nerveuse. A la lumière de ces expériences, c'est donc le mécanisme postulé par Licklider [22] qui paraît le plus vraisemblable. Licklider admet que la cochlée réalise une analyse en fréquence grossière, et que le système nerveux effectue une analyse temporelle d'autocorrélation sur les décharges nerveuses à l'intérieur de chaque canal de fréquence.

L'hypothèse de Licklider est plausible ; les neurones sont aptes à assurer les opérations nécessaires à un calcul d'autocorrélation : leur grand nombre permet d'imaginer un « autocorrélateur » nerveux pourvu d'assez de « lignes à retard » pour assurer une discrimination élevée des hauteurs sonores, en dépit de la faible résolution de l'analyse en fréquence réalisée par la cochlée. La surtension des résonateurs de la cochlée n'a plus alors de rapport direct avec la valeur des seuils différentiels en fréquence : correspond-t-elle à une donnée auditive ?

Dans l'hypothèse de Licklider, la détection de la hauteur se fait par une analyse d'autocorrélation à l'intérieur de chaque canal de fréquence. Trois sons sinusoïdaux, de fréquences respectives $f_1$, $f_2$ et $f_3$ non résolues par l'analyse cochléaire, pourront être

discriminés par l'analyse d'autocorrélation, s'ils sont présentés l'un aprés l'autre à l'oreille ; mais s'ils sont présentés simultanément, il y aura fusion, l'analyse d'autocorrélation portant sur la somme des décharges nerveuses correspondantes : cela conduira à un résidu. La surtension de la membrane basilaire doit donc être liée à l'espacement maximal des raies spectrales d'un résidu ; plus généralement, elle doit déterminer le pouvoir de résolution de l'oreille pour les sons simultanés : ce pouvoir de résolution doit être plus faible que pour les sons consécutifs. Plomp a étudié récemment le pouvoir analyseur de l'oreille [79] : il a établi que la loi d'Ohm [18], suivant laquelle l'oreille est capable de séparer une vibration complexe en ses composantes de Fourier, n'est valide que si l'écart de ces composantes est supérieur à une certaine valeur liée à la bande critique de l'oreille. Le concept de bande critique intervient dans plusieurs expériences sur l'audition, et tout d'abord dans celles de Fletcher cherchant à déterminer la largeur de bande d'un bruit intervenant effectivement pour masquer un son sinusoïdal [80] : les parties du spectre qui sont éloignées du son pur n'interviennent pas dans l'effet de masque. Ce concept intervient également dans l'additivité des intensités de composantes sonores [81]. La bande critique varie avec la fréquence, mais elle correspond sensiblement à une distance constante (1 mm) sur la membrane basilaire [82]. Les données de Plomp [79] confirment l'hypothèse de Greenwood [82] : la bande critique semble liée à la surtension de l'analyseur cochléaire.

Ces considérations nous conduisent à une explication cohérente du phénomène du résidu, tout en levant la contradiction entre l'amortissement de la membrane basilaire et la finesse de discrimination des hauteurs pour les sons graves. Le système cochléaire est très amorti : la membrane basilaire fonctionne comme un analyseur à large bande, dont la bande passante correspond approximativement à la bande critique ; c'est cette analyse qui détermine le pouvoir de résolution de composantes sinusoïdales simultanées et la fusion ou la possibilité de discrimination de ces composantes. La constante de temps correspondant à cette analyse est très faible, et l'intégration des données auditives est en général déterminée par les mécanismes nerveux intervenant après l'analyse : ces mécanismes peuvent donner lieu à plusieurs constantes de temps suivant les opérations réalisées. En particulier le mécanisme de détection de la hauteur détecte la périodicité à partir des maximums d'une sorte de fonction d'autocorrélation ; cela fait appel à une mémoire et c'est grâce à son intervention qu'est levé le paradoxe d'une sélectivité en mémoire élevée en dépit d'un amortissement important : l'analyse cochléaire n'a pas de rapport direct avec l'acuité de la perception des hauteurs. Le phénomène du résidu peut apparaître lorsque les composantes du son ne sont pas résolues par l'analyse cochléaire : la détection de périodicité est alors effectuée sur le complexe ; dans ce cas, l'information sur des hauteurs graves peut être transportée par des canaux nerveux qui correspondraient pour des sons purs à des fréquences élevées.

Le mécanisme que nous venons de décrire paraît cadrer avec nombre de résultats et donner une explication satisfaisante de la perception des hauteurs graves ou moyennes dans de nombreux cas. Toutefois il n'est certainement pas applicable à toutes les circonstances. Le fait, déjà mentionné, qu'on n'ait pu mettre en évidence nulle part dans la transmission nerveuse des fréquences supérieures à environ 4 000 Hz, conduit à admettre un mécanisme de localisation aux fréquences supérieures (on ne peut d'ailleurs obtenir de résidus à partir de composants de fréquence supérieure à 5 000 Hz [83]).

Pour concilier les diverses expériences, il semble qu'on doive admettre vers 400 Hz une transition floue entre détection de la périodicité (valant pour les fréquences inférieures) et analyse spatiale des fréquences (valant pour les fréquences supérieures), la position de cette transition dépendant de la nature du signal. Tonndorf [74] fait observer que la réponse spatio-temporelle de la membrane basilaire se rapproche davantage d'une analyse de Fourier dans le cas de sons sinusoïdaux et d'une analyse temporelle dans le cas de transitoires. Ward [84] estime que, pour les sons purs, la hauteur ne résulte de l'appréciation de la périodicité que pour des fréquences inférieures à 100 Hz, cependant que Flanagan et Guttman [10, 11, 12] montrent que pour des signaux formés de groupes d'impulsions quasi périodiques, la hauteur peut être déduite d'indices temporels jusqu'à 1 000 Hz ; ils font intervenir la résolution temporelle de la membrane basilaire pour interpréter leurs résultats. La différence entre indices spatiaux et temporels n'est pas toujours nette depuis que Békésy a décrit la progression de l'onde acoustique le long de la membrane basilaire : Békésy admet d'ailleurs [85], ainsi que Davis [86], que l'affinement spatial résultant de l'inhibition neurologique se fait grâce au décalage temporel entre l'excitation des différentes parties de la membrane basilaire. Pimonow [87] fait intervenir le regroupement des phases des influx nerveux comme facteur contribuant à la sélectivité auditive. Des mécanismes voisins sont invoqués par Davis [86] et Karplus [88]. La configuration spatio-temporelle de la membrane basilaire détermine des influx nerveux suivant des modalités complexes : ces influx sont interprétés en termes de hauteur de façon variable.

Le point de vue de Davis [86] sur la phylogenèse du sens auditif peut jeter quelque lumière sur la perception des hauteurs. Selon Davis, le sens auditif se réduisait primitivement à une sensibilité aux vibrations, qui ne permettait pas la détection et la discrimination des fréquences aiguës. (Békésy [85] a effectué plusieurs expériences rapprochant sens tactile et auditif.) L'apparition de la cochlée a fourni des mécanismes qui sont les seuls à jouer pour les

fréquences élevées, et qui se sont ajoutés aux mécanismes existants pour les fréquences graves et moyennes : à ces fréquences, les processus sont ainsi plus complexes et redondants, les discriminations plus fines et plus sûres. (C'est vraisemblablement pour cette raison que la plupart des identifications et différentiations de la parole ou de la musique se font dans ce domaine des fréquences basses ou moyennes.)

Pour terminer cet examen des théories de la perception des hauteurs sonores, nous ajouterons deux remarques. Tout d'abord, la plupart des manuels sur l'audition, se fondant sur des résultats, publiés en 1935, d'expériences portant sur un seul sujet [89], affirment que la hauteur d'un son simple de fréquence fixe varie beaucoup avec l'intensité du son. Mais des expériences récentes, accomplies avec plusieurs sujets, indiquent que cette variation est très faible pour la plupart des sujets, et qu'elle se fait dans différentes directions pour différents sujets [90, 91].

D'autre part, on arrive, dans certaines expériences élaborées, à manifester l'influence de l'audition binaurale sur la perception de la hauteur sonore — ce qui a conduit Licklider [23] à situer la perception de la périodicité à un niveau où l'interaction binaurale peut jouer un rôle. Mais ce rôle ne paraît important que dans des circonstances très particulières, comme celles réalisées dans l'expérience de Huggins [92].

### 2.2. Affinement dans la perception des hauteurs.

Nous avons mentionné, paragraphe 2.1, qu'il intervenait dans l'oreille interne des affinements spatiaux (ou spatio-temporels), sans lesquels la membrane basilaire ne jouerait qu'un rôle de prisme ne permettant pas une bonne résolution. Ces affinements peuvent tenir à des mécanismes complexes de l'organe de Corti [72] ou à une inhibition latérale [14] pouvant faire intervenir le décalage temporel entre l'excitation des divers points de la membrane basilaire [85, 86] ; ils conduisent à une distribution d'activité nerveuse plus sélective que la distribution d'activité mécanique.

Dans le cadre du modèle proposé par A. Moles (paragraphe 1.2), on peut envisager d'associer, à une analyse en fréquence assez floue, des mécanismes d'affinement : cela peut avoir une valeur d'illustration ; d'autre part ce type de mécanismes peut s'appliquer à l'analyse de spectres complexes, comme ceux qu'on rencontre par exemple en spectrométrie de masse [91] ou de résonance magnétique. Ainsi, pour la détection d'un maximum, un dispositif donnant la dérivée spatiale suivi d'un détecteur de zéro réaliserait un affinement. (On doit s'attendre à ce que les opérations effectuées par le système nerveux s'apparentent plus à des différences finies qu'à des dérivations exactes.) Nous avons écrit un programme donnant, à partir des deux ensembles d'échantillons, un coefficient de corrélation : à l'aide de ce programme, nous avons minimalisé la différence (au sens des moindres carrés) d'une courbe de forme donnée présentant un maximum, mais entachée d'incertitude, avec diverses courbes-modèles. Nous avons pu ainsi préciser la position du maximum avec une précision meilleure que celle qu'on peut atteindre par un ajustement local (on tient compte, en effet, de l'information sur toute la courbe).

Dans le cas d'un spectre complexe, on peut envisager plusieurs méthodes. Quand on ne sait rien *a priori* sur la structure du spectre, on peut, au moyen d'opérations d'autocorrélation, espérer faire ressortir des régularités. Dans d'autres cas, on pourra utiliser des méthodes de corrélation croisée entre le spectre considéré et des spectres « phénoménologiques » pour tenir compte de ce qu'on sait *a priori* du nombre et de la forme des raies : c'est ce que nous avons fait au moyen du programme évoqué ci-dessus, dans un cas simple. Un procédé semblable a été employé pour l'analyse d'une courbe de décroissance radioactive complexe : on a tiré parti de la connaissance qu'on avait des périodes radioactives des isotopes pour effectuer une première estimation des activités à l'origine et pour améliorer ensuite cette estimation [93].

Les méthodes que nous venons d'évoquer pour l'analyse des spectres reposent sur le principe fonctionnel suivant : les indications *a priori* sont mises à profit pour limiter l'arbitraire dans la synthèse de spectres « phénoménologiques », ces spectres résultant de l'addition des spectres des composants isolés agencés (par exemple pondérés) de façon variable ; cet agencement est ajusté pour minimaliser la différence (au sens d'un certain critère, comme celui des moindres carrés) entre spectre expérimental et spectre « phénoménologique » ; le spectre expérimental est assimilé au spectre « phénoménologique » le plus proche : on tire ainsi parti de tout le spectre, ce qui améliore la précision statistique. C'est là une « analyse par synthèse » : indépendamment de son intérêt technique [91, 95], cette modalité d'analyse nous permet d'interpréter les résultats d'une expérience sur l'analyse auditive d'un complexe harmonique (cf. 2.3*c*, 2.4 et 4.2.3).

### 2.3. Trois expériences sur la perception des hauteurs.

Pour éclaircir certains aspects de la perception des hauteurs, nous avons fait plusieurs expériences sommaires : trois d'entre elles nous paraissent mériter d'être décrites, pour leur intérêt de démonstration et aussi parce qu'elles peuvent susciter d'autres expériences plus poussées.

#### *a) Seuil et contexte.*

Nous avons présenté à deux sujets musiciens un certain nombre de couples de sons sinusoïdaux —

les sons d'un même couple ayant des fréquences légèrement différentes — en leur demandant d'indiquer pour chaque couple le son le plus haut. Les sons ont été engendrés à l'aide d'un calculateur électronique [96] : chaque son durait 1 s, avec des temps de montée et de descente de 50 ms ; les sons de chaque couple étaient séparés de 0,3 s. Nous n'avons retenu que les réponses pour un couple particulier $C_1$ comportant deux sons de fréquence $f_1 = 400$ Hz et $f_1' = 401$ Hz (l'écart $\delta f_1 = 1$ Hz rend la discrimination « difficile »). L'expérience a comporté deux phases, dans chacune desquelles $C_1$ était présenté dix fois (cinq fois dans l'ordre $f_1 - f_1'$, et cinq fois dans l'ordre inverse) en compagnie de vingt autres couples (dont six avaient une fréquence voisine de $f_1$ et quatorze des fréquences comprises entre 100 et 500 Hz) : l'ordre de présentation des couples avait été tiré au sort. Dans la phase $\alpha$, la différence de fréquence pour les couples du contexte était toujours supérieure à $\delta f_1$ ; dans la phase $\beta$ elle était toujours inférieure à $\delta f_1$ (et égale à zéro pour cinq couples). Les deux sujets ont subi les phases $\alpha$ et $\beta$ à deux jours d'intervalle ; le premier, qui les a subies dans l'ordre $\alpha$, $\beta$, a donné pour $C_1$ dix réponses correctes sur dix dans la phase $\alpha$ et huit sur dix dans la phase $\beta$ ; le second, qui les a subies dans l'ordre $\beta$, $\alpha$, a donné pour $C_1$ huit réponses correctes sur dix dans la phase $\alpha$ et cinq sur dix dans la phase $\beta$. Ces résultats indiquent que le contexte a influencé la discrimination, de façon apparemment paradoxale : le passage d'un contexte de sons plus faciles à discriminer à un contexte de sons moins faciles à discriminer n'a pas eu pour résultat une « adaptation » à une tâche plus difficile : tout se passe comme si les jugements faciles avaient orienté la discrimination. Selon Benzécri, il est possible que, dans le cas où les sons du contexte sont plus difficiles à discriminer, l'auditeur, pour tenter d'améliorer sa performance, fasse intervenir des mécanismes hiérarchiquement supérieurs, moins spontanés et plus conscients, mécanismes qui peuvent en fait troubler et amoindrir les facultés de discrimination : de même l'orthographe d'un mot qu'on écrit correctement « sans y penser » peut se mettre à poser des problèmes si l'on y réfléchit. Ce rôle du contexte confirme en tout cas que le seuil différentiel ne correspond pas à un pouvoir séparateur intrinsèque de l'oreille en tant que système physique.

Nous avons répété la phase $\beta$ pour le premier sujet et la phase $\alpha$ pour le second sujet, en leur demandant d'indiquer simplement s'ils estimaient, pour chaque couple, les sons identiques ou différents. Le premier sujet a estimé les sons du couple $C_1$ différents dix fois sur dix, et le second neuf fois sur dix. Les performances sont donc meilleures que l'orsqu'on demande aux sujets de préciser lequel des sons est le plus haut : cet effet correspond à celui étudié par Hirsch [56] pour la résolution temporelle de l'audition, mais il faudrait le confirmer par une étude plus élaborée.

### *b) Appréciation auditive de la périodicité.*

La hauteur d'un signal dont la périodicité rigoureuse est $T$, mais qui manifeste une quasipériodicité $T/2$ (Fig. 4), est-elle perçue comme correspondant à une sinusoïde de période $T$ ou de période $T/2$ ? Intuitivement, il doit y avoir une transition entre ces deux cas lorsque le rapport $\alpha$ entre les amplitudes de deux dents de scie successives (Fig. 4) décroît de 1 à 0.

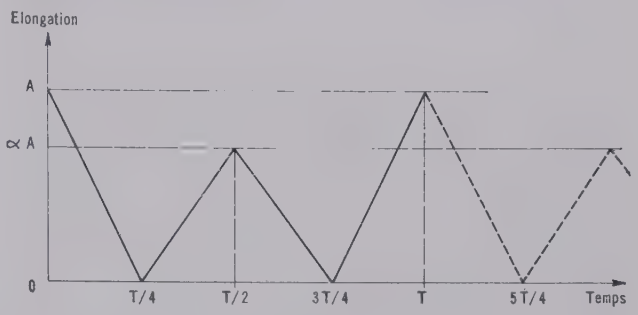


Fig. 4. — Onde à périodicité ambiguë, utilisée pour étudier la détection auditive de la périodicité.

A l'aide d'un calculateur électronique [96], nous avons engendré divers signaux de la forme indiquée figure 4, avec diverses valeurs de $\alpha$. Les sons correspondants ont été disposés par couples, le premier son de chaque couple étant un son de référence avec $\alpha = 1$ (périodicité rigoureuse $T/2$) et le second un son-test avec $\alpha \neq 1$. Chaque son avait une durée de 1 s, un temps de montée de 10 ms (qui ne crée pas de transitoire parasite [97]) et un temps de descente de 100 ms ; les sons de chaque couple étaient séparés de 0,5 s. Trois sujets musiciens, écoutant les couples produits, ont été d'accord pour conclure de la façon suivante :

— si $\alpha$ est très proche de 1 ($\alpha = 0{,}99$) la hauteur perçue est la même que celle du son rigoureusement périodique, de période $T/2$ et de fréquence $f = 2/T$ ;

— si $\alpha$ est très faible ($\alpha = 0{,}05$), la hauteur perçue se situe une octave plus bas (elle correspond à la fréquence $f/2$) ;

— entre ces deux cas extrêmes existe une région où l'on ne peut attribuer une hauteur unique au son-test, qui est perçu comme comportant une composante à la fréquence $f$ et une composante à la fréquence $f/2$.

Les résultats sont schématisés dans le tableau I :

TABLEAU I

| Fréquence fondamentale | Valeur de $\alpha$ | | | | |
|---|---|---|---|---|---|
| | 0,98 | 0,90 | 0,78 | 0,39 | 0,05 |
| $f = 200$ Hz | $f$ | $f \cup f/2$ | $f \cup f/2$ | $f \cup f/2$ | $f/2$ |
| $f = 100$ Hz | $f$ | $f \cup f/2$ | $f \cup f/2$ | $f \cup f/2$ | $f/2$ |
| $f = 50$ Hz | $f$ | $f$ | $f$ | $f \cup f/2$ | $f/2$ |

en fonction de la fréquence fondamentale $f$ du son de référence et de la valeur de $\alpha$, nous avons indiqué la fréquence correspondant à la hauteur perçue pour le son-test ; dans les cas intermédiaires, nous avons indiqué $f \cup f/2$, en soulignant le cas échéant la composante dominante.

Il est intéressant de comparer ces résultats à ceux de Flanagan, Guttman et Watson [98] qui ont fait une expérience semblable avec des signaux formés d'impulsions dont les amplitudes valent alternativement $A$ et $\alpha A$. Ces auteurs ont imposé à leurs sujets — non musiciens — d'accorder la fréquence d'un son sinusoïdal en faisant corespondre sa hauteur à celle du signal-test : il est probable que l'accord a été fait sur les composantes dominantes que nous avons indiquées ci-dessus. Avec cette hypothèse :

— l'oreille paraît moins sensible aux non-uniformités d'amplitude pour les sons graves : dans l'expérience citée [98], comme dans la nôtre, il faut une valeur de $\alpha$ plus faible pour provoquer la transition pour les fréquences plus basses. Cela se conçoit dans la mesure où l'attribut de hauteur correspondant à une périodicité de 25 Hz, s'il existe toujours [99], commence néanmoins à être difficile à dégager ;

— la comparaison des deux expériences indique que l'oreille est plus sensible aux non-uniformités d'amplitude pour les signaux que nous avons utilisés que pour des signaux impulsionnels. Ce résultat illustre la différence entre l'audition de signaux impulsionnels et de signaux dont les caractéristiques temporelles sont moins radicales.

Nous avons calculé le spectre de l'onde f($t$) indiquée figure 4 : la figure 5 représente les premières composantes du spectre pour diverses valeurs de $\alpha$. Entre certaines limites, les jugements sont indépendants de l'intensité : on ne peut interpréter les résultats en supposant que la hauteur correspond à la fréquence du premier harmonique ou à l'écartement des raies spectrales (en admettant dans ce dernier cas que l'oreille ignore les raies dont l'amplitude est en dessous d'un certain seuil). Il y a ici un conflit d'indices sensoriels conduisant à des hauteurs différentes : dans certains cas, aucun ne s'impose, d'où les ambiguïtés signalées dans le tableau I.

Outre les renseignements qu'ils peuvent apporter sur les processus de l'audition, les résultats de telles expériences sont intéressants pour la perception de la parole : parole pathologique (diplophonie) [100, 101] ou mécanisée. C'est ainsi que la moins grande sensibilité de l'oreille aux défauts d'amplitude pour des signaux se rapprochant d'impulsions périodiques confirme l'excellence du principe des vocoders à canaux codés où les composantes sont « en phase » (c'est le cas du vocoder construit au C. N. E. T. par G. Ferrieu).

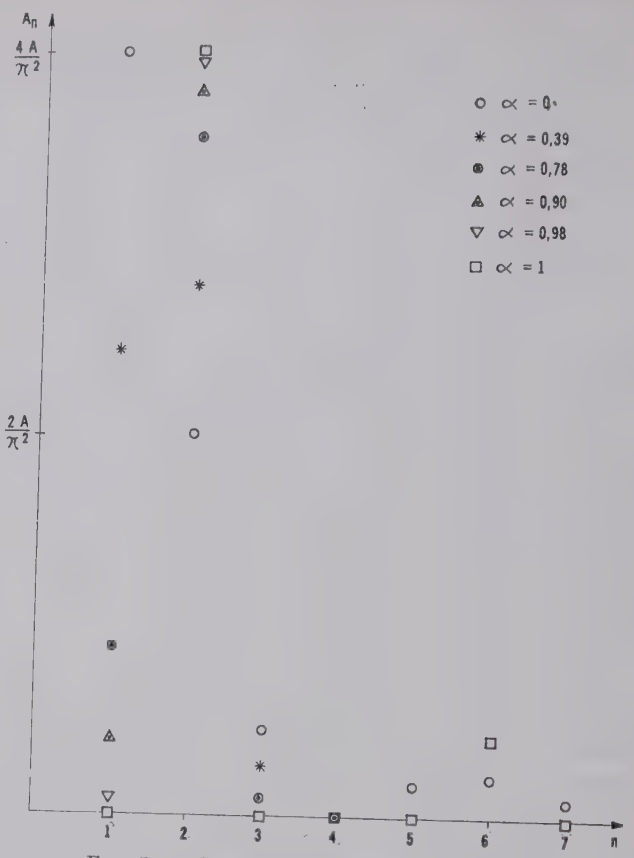


Fig. 5. — Spectres de l'onde de la figure 4 pour diverses valeurs de $\alpha$.

### *c) Analyse auditive d'un complexe harmonique.*

Nous avons demandé à cinq élèves musiciens, commençant leurs études d'écriture, d'analyser des accords que nous leur avons joués au piano, ou que nous leur avons situés dans les œuvres qu'ils connaissaient (par exemple : analyser le premier accord du premier scherzo de Chopin). Par analyser, nous entendions indiquer les notes musicales qui composaient l'accord (le « la » était donné avant chaque accord joué au piano). Nous avons remarqué que les élèves étaient incapables d'analyser certains accords au début de leurs études et qu'ils en devenaient capables après avoir étudié les chapitres du cours d'harmonie traitant de ces accords : plusieurs pouvaient alors faire de mémoire une analyse correcte (tel élève qui ne pouvait analyser le premier accord du premier scherzo de Chopin, l'a fait de mémoire trois mois après, sans l'avoir entendu entre temps ; dans certains cas, les analyses de mémoire étaient faites correctement à une transposition près).

Dans le cas de cette expérience, la détection des hauteurs ne se ramène pas à un processus passif d'analyse (fréquentielle ou temporelle) du signal acoustique reçu : on ne peut interpréter les résultats en supposant que c'est la résolution de la membrane basilaire qui limite les possibilités d'analyse, comme dans les expériences de Plomp [78], et les possibilités analytiques acquises sont en contradiction apparente avec les données de Pollack [103] selon lesquelles les observateurs seraient peu aptes à extraire d'un son complexe des composants non liés harmoniquement (même si les stimuli qu'il a utilisés sont trop différents des nôtres pour permettre une compa-

raison directe des résultats). Pour interpréter les résultats de cette expérience, il faut supposer que l'oreille opère une sorte d'« analyse par synthèse » (cf. 2.2).

### 2.4. Remarques sur la perception des hauteurs musicales.

La hauteur des sons est évidemment liée à leur fréquence : cependant, on s'est depuis longtemps rendu compte qu'il était plus adéquat de représenter la hauteur par une échelle en spirale ascendante que par une échelle linéaire [104], et qu'on pouvait décomposer la notion de hauteur sonore en deux attributs, l'un circulaire, le « chroma », qui est le même pour tous les sons en relation d'octave, et l'autre linéaire, la « hauteur brute », qui croît avec la fréquence. A ces attributs, on associe l'angle polaire et la cote d'un point de la spirale ; on peut remarquer l'analogie avec la mantisse et la caractéristique d'un logarithme. Cette conception est admise implicitement par les musiciens lorsqu'ils désignent par le même nom (la,..) des notes à intervalles d'octave, ou qu'ils parlent de monter une gamme ; elle a reçu une remarquable confirmation avec les expériences de Shepard, qui a synthétisé des stimuli pour lesquels les indices de hauteur brute sont noyés, par exemple des triplets de sons $A$, $B$, $C$, tels que $A$ soit perçu plus haut que $B$, $B$ plus haut que $C$, et $C$ plus haut que $A$ [105].

Depuis Pythagore, on a cherché à fonder les jugements relatifs aux hauteurs des sons musicaux sur des critères intrinsèques. Helmholtz [18] attribuait le sentiment de dissonance pour des sons simultanés (ou accords) à des battements entre ces sons : comme nous l'avons dit (cf. 1.3), cette théorie se heurte à de nombreuses difficultés. Plomp et Levelt [28] ont récemment proposé une théorie qui lie le caractère de dureté d'un accord et la relation entre bande critique de l'oreille et séparation en fréquence des partiels des sons de l'accord. Cette théorie paraît intéressante, mais incomplète, car elle ignore le rôle (souvent déterminant) du conditionnement culturel de l'auditeur ou du contexte musical. D'autre part, elle n'explique pas les jugements de rapports mélodiques : dans ce cas, les notes ne sont pas simultanées, mais successives, et les notes passées n'ont un effet que sous forme de « projection » nerveuse. La perception des rapports mélodiques fait également intervenir le conditionnement et le contexte musical. Ainsi, lorsque des violonistes professionnels jouent de la musique tonale, l'intervalle entre sensible et tonique [31] qui vaut théoriquement un demi-ton, est en fait systématiquement inférieur à sa valeur nominale : dans la musique tonale, la tonique a une fonction attractive sur la sensible, et l'intervalle perçu est influencé par cette fonction [106, et aussi 29, p. 73]. Dans un autre contexte, une telle déviation de l'intervalle nominal est perçue comme fausse.

On pourrait citer bien d'autres exemples qui montrent que la hauteur perçue ne résulte pas d'une donnée brute des sens. C'est dans le domaine de la musique, où la perception de la hauteur est très sollicitée, que les exemples sont les plus frappants (cf. § 2.3*c*). Cela nous conduit à considérer la hauteur comme un attribut déduit des données sensorielles par un processus actif (bien que plus ou moins inconscient). La membrane basilaire donne divers indices correspondant à des mouvements aussi bien qu'à des configurations : à partir d'une onde $s(t)$, le système nerveux transmet :

$$r(t) = p[(s(t)],$$

où $p$ est une anamorphose complexe : c'est de $r(t)$ qu'est extrait l'attribut de hauteur. Cette extraction se fait suivant des modes variables suivant la nature du signal (cf. § 2.1, [107]) et aussi suivant l'état de l'individu [30], son conditionnement (cf. § 2.3*c*), la façon dont il interprète le contexte (cf. § 2.3*a* [34]). Un tel processus nous paraît seul capable de concilier les nombreuses données psychoacoustiques ; il est plausible du point de vue physiologique, puisque, selon Mary Brazier et ses collaborateurs [108], la réponse cérébrale aux stimulations sensorielles dépend plus des événements — passés ou présents — à l'intérieur du cerveau, que des paramètres du stimulus. Le cerveau effectue donc des opérations pour concilier les indices, d'origine externe ou interne, relatifs à la hauteur. Le conditionnement peut aider ces opérations en permettant de tester des hypothèses antérieurement rencontrées : c'est le cas de l'expérience décrite paragraphe 2.3*c* ; le processus compare alors à $r(t)$ divers $p[s_i(t)]$ internes et assimile $s(t)$ à un $s_i(t)$ particulier : on peut parler alors d'analyse par synthèse. Parfois, un compromis satisfaisant entre indices contradictoires n'est pas possible (cf. § 2.3*b*) : l'oreille exceptionnelle de Varèse pouvait dans certaines circonstances se refuser à attribuer une hauteur unique à un son périodique. Par contre, dans de nombreux cas, une hauteur unique est attribuée à un son dont la fréquence peut comporter une marge importante de fluctuations ou de variations : c'est le cas pour les notes chantées [29] ; nous avons pu vérifier que c'était aussi le cas pour les sons dont la partie initiale comporte un glissando (variation monotone de fréquence) assez rapide et s'étendant sur une octave, en synthétisant de tels sons à l'aide d'un calculateur électronique [93].

De façon générale, la musique joue sur les hauteurs des sons, entre lesquelles elle organise des relations élaborées. Il nous paraît adéquat de définir un son musical comme un son donnant lieu à des indices sensoriels d'où l'on puisse dégager une hauteur — sous réserve que cet attribut de hauteur soit assez bien défini pour se prêter à l'établissement de relations cohérentes avec d'autres sons. Cette réserve refuse le qualificatif de musical à des sons de fréquence fondamentale supérieure à 5 000 Hz environ [109], qui ne sont d'ailleurs presque jamais utilisés en musique, et aussi à des sons trop graves

(la « fréquence de coupure » dépend de la façon dont on définit l'établissement de relations cohérentes avec d'autres sons, mais elle se situe aux alentours de 60 Hz [99] ; des sons plus graves sont utilisés en musique, mais rarement, et plus pour colorer que pour organiser le discours musical). Le son musical n'est nullement assimilé au son périodique : aucun son n'est d'ailleurs rigoureusement périodique ; mais des sons apériodiques comme les sons graves du piano sont incontestablement des sons musicaux.

La hauteur des sons musicaux se prête à l'établissement de relations élaborées et cohérentes ; cette possibilité est propre au chroma, plus qu'à la hauteur brute : cela ressort clairement de la théorie musicale [31]. Ce sont les différences de hauteur brute qui font paraître une même note « plus aiguë », « plus haute » lorsqu'elle est chantée par un ténor que par un contralto : la hauteur brute est un aspect du timbre ; les propriétés de la hauteur brute dans le domaine des sons rapprochent cet attribut de la couleur dans le domaine visuel. En restant sur le plan des hypothèses, il nous paraît que le chroma correspond surtout à des indices temporels, alors que la hauteur brute traduit la distribution d'excitation le long de la membrane basilaire (et aussi sans doute nombre d'indices moins bien connus, comme le sentiment d'effort plus grand pour le ténor que pour le contralto dans l'exemple précédent) : cette hypothèse expliquerait la force de la relation d'octave (surtout si l'on admet le processus exposé paragraphe 2.1, où l'analyse d'autocorrélation est exercée après une analyse en fréquence grossière) et à un moindre degré des relations de quinte, de quarte,... On voit mal comment une hauteur inférée à partir de la distribution d'excitation de la membrane basilaire se prêterait à des relations ou des comparaisons fines comme celles qui interviennent en musique ; d'ailleurs, les fréquences supérieures à 4 000 Hz ne sont pas des hauteurs musicales, les intervalles musicaux courants n'y sont plus reconnaissables [109], il ne subsiste plus de chroma [110] : or, à ces fréquences, le seul indice qui subsiste pour la hauteur est la distribution d'excitation de la membrane basilaire.

Un musicien doit estimer des relations de hauteur quantifiées : au cours de son éducation ou de son conditionnement, il élabore (en général inconsciemment) des stratégies pour juger en fonction de ces relations. La quantification des échelles de fréquence (gammes) et la représentation symbolique qui lui est liée lui permettent de pallier des incertitudes, et d'arriver à des jugements sûrs et reproductibles, mais fonctions de son conditionnement. Nous avons pu vérifier que des musiciens exercés à la seule musique tonale faisaient plus d'erreurs, dans des dictées musicales tonales comportant des notes étrangères à la tonalité, que des musiciens ayant une certaine connaissance de la musique atonale. Nous avons pu vérifier également que les musiciens occidentaux notent souvent les musiques exotiques en assimilant les hauteurs de la gamme exotique aux échelons les plus proches de la gamme chromatique occidentale. L'expérience de Meyer [34] relève du même processus.

L'éducation du musicien a ainsi développé en lui un réseau de références et des stratégies supplémentaires, particulièrement de tests d'hypothèse, d'analyse par synthèse, en relation avec l'« écoute intérieure ». Cela explique l'observation ci-dessus sur les dictées musicales : les sujets habitués à la musique atonale sont plus aptes à inclure dans les « hypothèses » qu'ils testent des éléments étrangers à un contexte tonal ; cela explique aussi l'expérience que nous avons citée (cf. § 2.3*c*) qui contredit apparemment les expériences de Plomp [79] et Pollack [103] effectuées dans un contexte non musical ; cela explique enfin que des sujets musicaux puissent avoir des possibilités d'identification sonore considérablement supérieures [50] à celles de sujets non musiciens [48]. L'aptitude qu'on appelle oreille absolue [110] nous paraît d'ailleurs acquise plutôt qu'innée : le fait que cette aptitude soit rencontrée bien plus fréquemment dans les familles de musiciens peut tenir aussi bien au contexte culturel qu'à l'hérédité.

Ce point de vue, suivant lequel apparaissent avec l'éducation des processus de plus en plus élaborés de discrimination des hauteurs, est confirmé par une observation récente. Tanner et Rivette [111] ont remarqué que trois sujets, pourvus de bonnes aptitudes à discriminer les intensités sonores, étaient pour ainsi dire sourds aux hauteurs sonores, pratiquement incapables de distinguer des sons par leurs différences de hauteur : la langue maternelle de ces trois sujets était une langue indienne, le punjabi. Il semble que ces individus, possédant en puissance les mécanismes nécessaires, n'aient pas développé leurs possibilités de discrimination des hauteurs, parce que le besoin n'en existait pas dans leur langue maternelle (les possibilités de développement de certains mécanismes peuvent dépendre crucialement de l'âge du sujet).

Nous allons à présent nous pencher sur des mécanismes de corrélation et d'autocorrélation, invoqués par A. Moles [2] et Licklider [22] pour leurs modèles auditifs.

## 3. CORRÉLATION, AUTOCORRÉLATION ET AUDITION.

### 3.1. Cas d'un phénomène stationnaire.

Soit une fonction $f(t)$ décrivant un processus stationnaire ayant un contenu de puissance (par exemple une vibration périodique, un bruit à bande limité, un débit continu de parole). La fonction d'autocorrélation de $f(t)$ est définie comme :

$$\Phi(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t)\, f(t+\tau)\, dt.$$

On peut montrer que la fonction d'autocorrélation est maximale pour $\tau = 0$ ; qu'elle est fonction paire de $\tau$ ; qu'elle élimine l'information concernant les phases respectives des harmoniques d'un signal

périodique ; que plus précisément (théorème de Wiener) elle est la transformée de Fourier du spectre de puissance :

$$S(\omega) = \lim_{T\to\infty} \frac{1}{T} \left| \int_{-T}^{T} f(t)\, e^{-j\omega t}\, dt \right|^2, \quad [112].$$

Le processus d'autocorrélation est exploité pour la détection des signaux périodiques dans le bruit : la fonction d'autocorrélation d'une fonction périodique est une fonction périodique de même période, alors que celles de fonctions aléatoires représentatives de nombreux bruits tendent vers zéro si $\tau$ tend vers l'infini ; cela permet la séparation des composantes périodiques et aléatoires d'autant plus commodément qu'un autocorrélateur est plus facile à réaliser qu'un filtre à bande très étroite [109].

La fonction de corrélation des fonctions $f(t)$ et $g(t)$ est définie comme :

$$\Psi_{fg}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t)\, g(t+\tau)\, dt.$$

Le processus de corrélation est également utilisé pour la détection des signaux périodiques dans le bruit : on calcule la corrélation du message reçu avec une fonction modèle qui décrit le signal à détecter en l'absence de bruit.

### 3.2. Cas d'un phénomène variable au cours du temps.

L'approximation accomplie en supposant le phénomène stationnaire et le temps d'intégration infini ne sera valable que si l'on choisit un temps d'intégration grand par rapport aux périodes des fluctuations les plus lentes. Or l'oreille intègre avec de faibles constantes de temps (cf. § 1.4*c*) et, à la différence de ce qui se passe en optique, les phénomènes transitoires et les variations temporelles jouent un grand rôle en acoustique. Il sera donc utile de restreindre le calcul des fonctions d'autocorrélation ou de corrélation à des sections du processus complet, de façon à ce qu'on puisse suivre la variation dans le temps de ces fonctions. D'ailleurs les fonctions définies plus haut n'ont plus de sens pour des phénomènes non stationnaires, puisqu'elles supposent une intégration sur l'avenir.

Fano [113] a étudié les propriétés d'une fonction d'autocorrélation définie par :

$$\varphi_t(\tau) = 2\alpha \int_{-\infty}^{t} f(x)\, f(x-\tau)\, e^{-2\alpha(t-x)}\, dx,$$

où la « fonction de poids » (ou « fonction de fuite ») $e^{-2\alpha(t-x)}$ tend à diminuer la contribution à l'intégrale des composantes éloignées dans le temps. Une telle « fuite » joue un rôle comparable à la mémoire phosphorescente du modèle proposé par A. Moles (§ 1.2*f*). On peut réaliser un corrélateur analogique correspondant à cette définition (Fig. 6*a*) : avec la fonction de poids : $p(t) = e^{-2\alpha(t-x)}$, choisie par Fano, le filtre se réduit à un intégrateur RC de constante de temps $1/2\,\alpha$. Un théorème de Wiener modifié s'applique à cette définition de la fonction d'autocorrélation, la reliant à un spectre de puissance à court terme (ou spectre instantané) convenablement choisi [113].

On peut choisir pour fonction de poids $p(t)$ une fenêtre rectangulaire : cela correspond à la mémoire phosphorescente symbolisée paragraphe 1.2*f* par une autocorrélation portant sur un certain nombre de cartes. La fonction d'autocorrélation correspondante est :

$$\varphi_t(\tau) = \int_{t-(T-\tau)}^{t} f(x)\, f(x-\tau)\, dx,$$

et cette fonction est la transformée de Fourier d'un spectre :

$$G(\omega, t) = \left| \int_{t-T}^{t} f(x)\, e^{-j\omega x}\, dx \right|^2.$$

Le calcul analogique de ce spectre fait appel à des filtres de réponse impulsionnelle $\cos \omega t$ et $\sin \omega t$, irréalisables mais simulables sur calculateur.

Schroeder et Atal [114] ont donné des définitions de fonctions d'autocorrélation et de spectres dépendant du temps assez générales, en spécifiant comment les mesures correspondantes pouvaient être réalisées physiquement ; ils ont modifié le théorème de Wiener liant fonction d'autocorrélation et spectre en puissance, pour l'étendre aux cas où les mesures sont faites avec des filtres physiquement réalisables. Ils définissent la fonction d'autocorrélation par :

$$\varphi(\tau, t) = \int_{-\infty}^{t} f(x)\, f(x-|\tau|)\, p(t-x)\, dx,$$

où $p(t)$ est une fonction de poids arbitraire, mais physiquement réalisable (en particulier $p(t) = 0$ pour $t < 0$). Cette fonction peut être calculée par un circuit schématisé figure 6*a*. D'autre part, le spectre en puissance à court terme est défini par :

$$W(\omega, t) = 2 \int_{0}^{\infty} m(\tau) \cos \omega\tau\, d\tau \times \int_{-\infty}^{t} f(x)\, f(x-\tau)\, p(t-x)\, dx ;$$

soit encore :

$$W(\omega, t) = 2 \int_{-\infty}^{t} f(x)\, p(t-x)\, dx \times \int_{0}^{\infty} m(\tau)\, f(x-\tau) \cos \omega\tau\, d\tau$$

Ce spectre est donné par un circuit schématisé figure 6*b*. Avec ces définitions, on a les relations :

$$W(\omega, t) = \int_{-\infty}^{+\infty} \varphi(\tau, t)\, m(|\tau|) \cos \omega\tau\, d\tau,$$

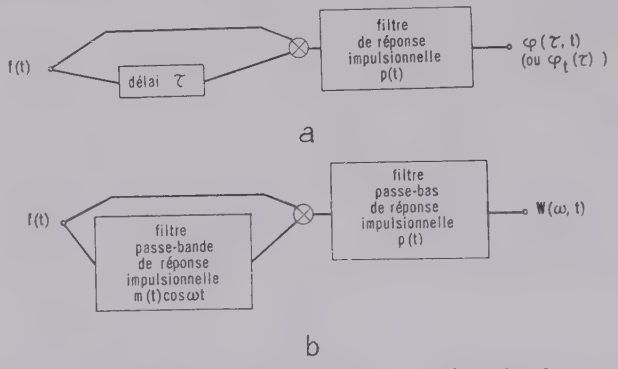


FIG. 6. — Circuits analogiques donnant une fonction d'autocorrélation à court terme (*a*) et un spectre de puissance à court terme (*b*).

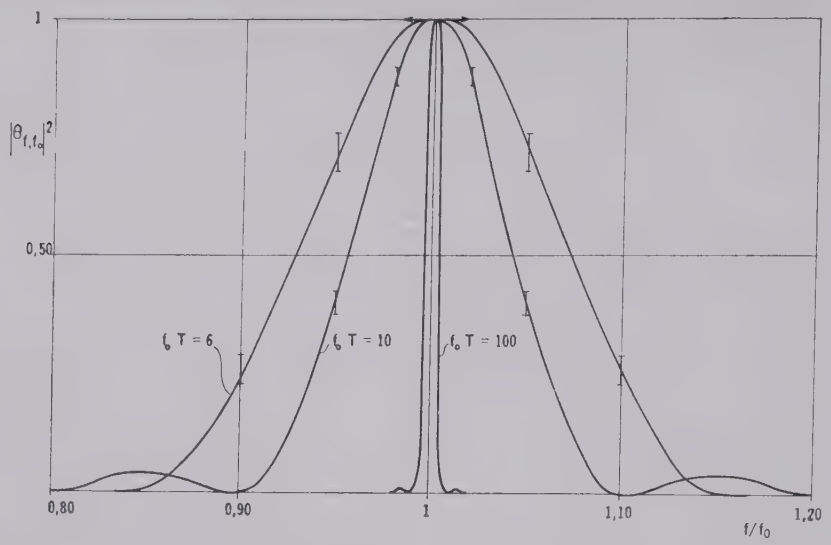


FIG. 7. — Courbes de sélectivité, donnant $|\theta_{f,f_0}|^2$ en fonction de $f/f_0$ pour trois valeurs de $f_0\,T$.

et :

$$\varphi(\tau, t) = \frac{1}{2\pi m(|\tau|)} \int_{-\infty}^{+\infty} W(\omega, t) \cos \omega\tau \, d\omega.$$

Ainsi spectre de puissance et autocorrélation donnent les mêmes informations ; cependant, en raison de la nature des opérations effectuées (moyennes pondérées réduisant la définition spectrale), la fonction d'autocorrélation donne des informations plus précises sur le spectre en puissance que la mesure directe du spectre en puissance avec des filtres et la même fonction de poids.

### 3.3. Une méthode d'analyse en fréquence réalisable sur calculateur.

La relation entre autocorrélation et filtrage a été mise à profit par Blackman et Tukey [115] pour définir une méthode d'analyse en fréquence : cette méthode, qui consiste à prendre la transformée de Fourier de la fonction d'autocorrélation, est couramment mise en œuvre sur calculateur. L'analogie entre filtrage et corrélation nous a suggéré une autre méthode : intuitivement la corrélation entre une fonction du temps $F(t)$ et la fonction $e^{2\pi i f_0 t}$ doit jouer un rôle analogue à celui d'un filtre de fréquence centrale $f_0$. A quoi correspond, pour la corrélation, la sélectivité du filtre ?

Soit :

$$_F\theta_{f_0} = \frac{1}{T} \int_0^T F(t) \, e^{2\pi i f_0 t} \, dt.$$

Comme dans le cas d'un filtre, nous étudierons la sélectivité de $|_F\theta_{f_0}|^2$ en utilisant pour « fonction d'entrée» $F(t)$ une sinusoïde $A \sin(2\pi f t + \varphi)$. Posons :

$$\theta_{f,f_0} = \frac{1}{T} \int_0^T A \sin(2\pi f t + \varphi) \, e^{2\pi i f_0 t} \, dt.$$

Dans la suite nous supposerons qu'on choisit un ntervalle d'intégration $T$ multiple de $1/f_0$, de façon que $f_0\,T$ soit entier. Alors, si $f \to 0$, $|\theta_{f,f_0}|^2 \to 0$ ; si $f \to \infty$, $|\theta_{f,f_0}|^2 \to 0$ ; si $f \to f_0$, $|\theta_{f,f_0}|^2 \to A^2/4$. Le valeur de $|\theta_{f,f_0}|^2$ dépend de la phase, mais cette dépendance est de moins en moins sensible quand la valeur de $f_0\,T$ augmente. Sur la figure 7, on a porté les courbes donnant $|\theta_{f,f_0}|^2$ en fonction de $f/f_0$ pour $f_0\,T = 6$, $f_0\,T = 10$, $f_0\,T = 100$ (en supposant $A = 2$) ; les intervalles figurés avec les points portés représentent la latitude de variation de $|\theta_{f,f_0}|^2$ quand la phase $\varphi$ varie. On voit sur cette figure que la sélectivité est d'autant plus grande que $f_0\,T$ est plus élevé (c'est-à-dire que $T$ est plus élevé pour une fréquence centrale $f_0$ donnée). Cette conclusion illustre la contradiction entre sélectivité et brièveté d'analyse et le rôle de la mémoire dans l'augmentation de la sélectivité (cf. 2.1).

Étant donnée une fonction $F(t)$, on peut donc en réaliser une analyse spectrale en calculant $|_F\theta_{f_0}|^2$ pour diverses valeurs de $f_0$ : cela correspond à une analyse par une batterie de filtres dont les fréquences centrales seraient les valeurs attribuées à $f_0$. La sélectivité croît avec la valeur de $f_0\,T$.

Cette méthode est facile à mettre en œuvre sur calculateur : on calculera :

$$\left| \frac{1}{T} \int_0^T F(t) \, e^{2\pi i f_0 t} \, dt \right|^2,$$

à partir des valeurs de la fonction $F(t)$ par une méthode classique d'approximation des intégrales définies. En augmentant la valeur de $T$ choisie, on augmentera la sélectivité de l'analyse, mais aussi de temps de calcul. D'autre part, pour une fonction $F(t)$ non stationnaire, on pourra effectuer plusieurs calculs successifs des $|_F\theta_{f_0}|^2$ pour obtenir l'évolution temporelle du spectre. La méthode est avantageuse pour faire une analyse plus détaillée dans certaines régions du spectre : dans ce cas, la méthode de Blackman et Tukey [115] impose le calcul de la fonction d'autocorrélation avec la précision maximale requise.

### 3.4. Autocorrélation et corrélation dans l'audition.

Nous avons vu plus haut (cf. § 2.1) que la théorie de Licklider pour la perception des hauteurs [22] supposait des analyses d'autocorrélation. Le système nerveux paraît approprié à réaliser des opérations du type corrélation : une chaîne de neurones est comparable à une ligne à retard, et la multiplication ou l'intégration s'y conçoivent bien [22]. L'audition semble indiquer des mécanismes proches de la corrélation ou de l'autocorrélation dans d'autres processus que celui de la perception des hauteurs [1].

#### *a) Appréciation de la « coloration » et spectres à court terme.*

Un bruit est « coloré » si son spectre n'est pas uniforme. Il semble que, pour apprécier la coloration du timbre, l'oreille fasse appel à des mécanismes temporels du type autocorrélation, appréciant non pas le spectre lui-même, mais un spectre à court terme [113]. Cette hypothèse a été confirmée par des expériences qui ont permis de préciser le type de « fonction de poids » (cf. § 3.2) qui intervenait dans l'audition : c'est une fonction monotone décroissante dont la pente initiale correspond à une constante de temps de 9 ms [116]. Mais surtout, cette hypothèse semble la seule qui donne une explication satisfaisante de l'insensibilité de l'oreille aux irrégularités de la courbe de réponse en fréquence d'une salle [116]. En effet, dans une salle de concerts ou de conférences jugée comme ayant une bonne acoustique, la transmission varie très irrégulièrement avec la fréquence : les irrégularités, de l'ordre de 10 dB, dépassent souvent 30 dB [117] ; or ces irrégularités ne sont pas gênantes subjectivement. Dans la mesure où l'oreille ne perçoit que les irrégularités du spectre à court terme, sa résolution en fréquence sera insuffisante pour apprécier les irrégularités intrinsèques.

Si l'on admet que l'oreille perçoit un spectre à court terme, la question de sa sensibilité à la phase en est clarifiée. La loi d'Ohm, qui stipule que l'oreille est insensible à la phase [18], est manifestement insuffisante : dans les lignes de téléphone à grande distance, on égalise les phases pour éviter que la parole ne soit trop déformée et rendue inintelligible. Pourtant, dans bien des cas, la phase a peu d'influence sur le timbre des sons. Ainsi il est difficile de modifier le timbre d'un son stationnaire proche d'une voyelle en ne modifiant que les conditions de phase [118] ; si l'on analyse le son en signaux élémentaires de durée efficace 20 ms, on peut changer les phases des composantes sans que l'oreille s'en aperçoive en général [54] Mais dans les cas où les distorsions de phase correspondent à des retards de plus de 50 ms, le spectre à court terme sera modifié, et cette modification sera audible — comme une sorte de traînage, de réverbération.

L'oreille peut d'ailleurs être sensible à la phase dans d'autres cas ; il est possible de « jouer un air » en ne faisant varier, dans un son complexe soutenu, que les phases de certains harmoniques [119]. Pour que des variations de phase soient audibles, il semble nécessaire qu'elles se traduisent par des modifications sensibles de l'enveloppe de la forme d'onde temporelle (par exemple, si la forme d'onde se rapproche d'une impulsion périodique, des modifications de phase risquent de modifier cette forme et le timbre perçu) [120]. D'autres conditions sont plus discutées [121]. Mais il importe de noter que ces effets de phase risquent d'être noyés dans une salle réverbérante, où se produisent des interférences aléatoires : les expériences citées [120, 121] ont été faites avec des écouteurs, et il ne semble pas que l'on ait observé d'effets de phase appréciables dans un environnement réverbérant, sauf s'ils correspondent à une modification du spectre à court terme. Il serait toutefois utile de reprendre plusieurs expériences accomplies sur la perception de la phase, en contrôlant et en précisant rigoureusement les conditions de réverbération (écoute avec un ou deux écouteurs, un ou deux haut-parleurs, en salle sourde ou réverbérante).

#### *b) Corrélation, filtrage, éducation et reconnaissance.*

La perception humaine est apte à reconnaître des configurations (par exemple, des lettres manuscrites ou des mots parlés). Nous verrons dans le chapitre suivant, consacré à la reconnaissance des formes, que la corrélation avec des configurations de référence est couramment utilisée pour la reconnaissance automatique des configurations.

Diverses données psychologiques ont conduit Hebb à proposer l'hypothèse suivante [122] : chez l'homme, l'éducation, le conditionnement, l'apprentissage sensoriel se traduiraient par la constitution de réseaux cérébraux ; ces réseaux « filtreraient » sélectivement les données sensorielles, ce qui conduirait à la reconnaissance des configurations. L'analogie entre filtrage et corrélation suggère que les reconnaissances pourraient encore faire appel à la corrélation des données sensorielles avec des configurations de référence emmagasinées dans la mémoire — de la même façon qu'on peut identifier une clé, soit en l'essayant à la serrure, soit en la comparant à un double. Entre le filtrage d'une fonction $f$ par un réseau $G$, et la mise en corrélation de $f$ avec une fonction modèle $g$, existe la différence entre processus passif et actif : il faut évoquer la fonction modèle $g$. On pourrait donc songer à rattacher à la corrélation les processus actifs de reconnaissance, et au filtrage des processus de perception moins élaborés, de type plus végétatif (cf. § 2.3 et 4.2.4). Cependant, la différence entre filtrage et corrélation s'estompe quand on passe des filtres linéaires aux possibilités non linéaires présentes dans le système nerveux ou la machine numérique : nous avons vu (cf. § 3.2) que la corrélation avec la fonction $\cos \omega t$ se ramenait au filtrage par un réseau de

réponse impulsionnelle cos $\omega t$, simulable sur une machine numérique. On a émis l'hypothèse que le rôle de l'attention, la raison d'être des titres, annonces ou indicatifs, était d'exciter, de « mettre sous tension » sélectivement certains réseaux cérébraux, condition nécessaire pour qu'ils puissent opérer [123]. En raison de la nature des opérations nerveuses, une simulation sur calculateur d'un processus cérébral schématisé par la corrélation ou le filtrage paraît plus appropriée qu'une simulation analogique. On remarquera cependant qu'il est incorrect de considérer le codage nerveux comme purement numérique, et que la cellule nerveuse isolée ne paraît pas agir seulement par tout ou rien [124].

Il faut noter que l'hypothèse séduisante de réseaux nerveux, qui explique bien certains comportements [122, 123], n'a pu être confirmée physiologiquement ; on pense maintenant que le support matériel de la mémoire à long terme est chimique et fait intervenir les molécules d'ARN ou peut-être d'ADN [125]. Mais surtout, les opérations effectuées dans le cerveau ne correspondent sans doute pas exactement à des opérations mathématiques simples comme l'autocorrélation. Dans le chapitre suivant, nous verrons la distance qui sépare la reconnaissance artificielle des configurations et les processus humains de reconnaissance des formes, processus plus riches et plus complexes.

## 4. RECONNAISSANCE DES FORMES. CONCLUSION.

Nous avons vu paragraphe 1.3 qu'une simulation de l'audition, ne se réduisant pas à l'illustration de mécanismes physiques périphériques, doit opérer des reconnaissances de formes, plutôt que des analyses physiques brutes.

Un modèle de l'audition tel que celui proposé par Moles (cf. § 1.2) peut être envisagé sous un autre point de vue : on peut en partir pour réaliser un dispositif de traitement d'information sonore. On peut tenter d'utiliser les mécanismes d'intégration qu'il comporte pour assurer le dépouillement d'électrocardiogrammes ou d'enregistrements sismiques, ou la reconnaissance automatique d'un vocabulaire parlé. C'est ainsi que Martin et ses collaborateurs [126] mettent au point un système imitant l'audition humaine et visant à reconnaître les phonèmes anglais. Dans une telle perspective, il faut aménager le modèle pour la meilleure solution du problème technique que l'on se pose. Aussi, dans la première partie, paragraphe 4.1, de ce chapitre, nous étudierons la reconnaissance automatique des configurations par des systèmes artificiels : après avoir indiqué des conditions à remplir, nous donnerons quelques résultats nouveaux sur les processus optimaux dans certains cas. (Pour les démonstrations, cf. [1].)

Dans la seconde partie, paragraphe 4.2, nous aborderons la question de la reconnaissance des formes dans la perception humaine. Plusieurs auteurs ont cru pouvoir étendre à ce domaine des considérations valables dans le domaine technique : à propos de la parole et de la musique, nous insisterons plutôt sur les différences entre ces deux domaines.

### 4.1. Cas des systèmes artificiels : modèles pour machines.

#### *4.1.1. Conditions nécessaires.*

Dans le domaine technique de la reconnaissance automatique, on considère généralement le problème de la reconnaissance des formes d'une manière restrictive : reconnaître une configuration (nous dirons par la suite, plus brièvement, une forme) dans un phénomène physique, c'est décider de son appartenance à l'une ou l'autre de plusieurs classes. Par exemple, reconnaître une lettre dans un message lumineux, c'est décider que ce message appartient à la classe de tous les messages représentant cette lettre. De même, un mot du langage parlé correspond à une forme : différents locuteurs prononçant ce mot émettront des signaux physiquement différents, mais l'auditeur pourra y reconnaître le mot [103].

L'opération de reconnaissance des formes suppose un certain nombre d'indications, de tests ou de mesures sur le phénomène physique, et une décision inférée à partir de ces indications, du résultat de ces tests ou de ces mesures. On peut donc y distinguer les deux stades suivants [127].

— Réception. Cette phase correspond au recueil des informations (et éventuellement à leur traitement, préalablement à la décision) : elle définit un espace $E$ de paramètres de reconnaissance ; à chaque phénomène sera associé un point dans cet espace.

— Catégorisation. A chaque point de $E$, il faut associer l'une des classes. La stratégie de reconnaissance définit donc une partition de l'espace $E$ en cellules : les points d'une même cellule correspondent aux phénomènes d'une même classe. La décision — la reconnaissance proprement dite — consiste à reconnaître l'appartenance du point de $E$ représentatif du phénomène à l'une ou l'autre de ces cellules.

La reconnaissance des formes se distingue d'une mesure physique brute en particulier en ce que le processus de reconnaissance ne doit pas être affecté par certains groupes de transformations ; par exemple, l'identification d'une lettre doit être indépendante de changements de taille ou de position : or, si la « réception » est accomplie par une mosaïque de cellules photoélectriques, la même lettre, après avoir subi un déplacement, n'impressionnera pas les mêmes cellules. Il faut donc que, par rapport à l'espace physique des phénomènes, l'espace des configurations ait une structure d'ensemble — quotient par la relation d'équivalence suivante : deux phénomènes sont équivalents s'ils se correspondent par des transformations du groupe ; ou en tout cas,

il faut que chaque cellule soit globalement invariante par rapport au groupe transmué agissant sur les points de l'espace $E$. Dans la référence [103], nous donnons un exemple de processus d'identification dans un espace à deux dimensions échantillonné (permettant de représenter une forme visuelle ou une courbe à deux dimensions), processus qui n'est pas altéré par des translations de la configuration.

Il existe théoriquement une méthode générale pour assurer l'invariance de la reconnaissance par rapport à un groupe de transformations. Le bruit des canaux de transmission introduit des seuils physiques, ce que traduit la formule Shannon-Tuller donnant la capacité informationnelle d'un canal physique [128] ; on peut donc considérer que le groupe de transformations est fini. Si la réception associe à une configuration $F_i$, un paramètre $f(F_i)$, associons à une figure donnée $F_k$ l'ensemble $F_k = F_u$, $F_v$, ... de toutes les figures équivalentes à $F$ par l'ensemble des transformations du groupe. La moyenne $f(F_k)$ des $f(F_u)$, $f(F_v)$, ... a une valeur indépendante du choix de $F_k$ parmi une classe d'équivalence définie par les transformations. Une reconnaissance effectuée à partir de telles moyennes sera donc invariante par toute transformation du groupe. Cependant, si le nombre de transformations considérées est fini, il est en général très grand et cette méthode conduit à des procédés qui ne sont applicables pratiquement qu'à des cas très limités (toutefois, on a soutenu l'hypothèse qu'ils étaient réalisables par le cerveau [129]).

La reconnaissance des formes doit encore satisfaire à une autre condition : deux phénomènes très « proches » doivent « presque toujours » appartenir à la même classe. Cela suppose la définition dans l'espace $E$ d'une topologie, au sens de laquelle chaque cellule devra être un ouvert. (Ce n'est pas le cas si par exemple on divise la droite réelle en deux classes, celle des points à coordonnées rationnelles et celle des points à coordonnées irrationnelles : mais un tel classement n'a pas de sens physique.) Le plus souvent, on définira dans $E$ une métrique, qui peut intervenir dans la partition de l'espace. Par exemple, la partition peut être effectuée par des sphères dont chacune entoure un point de référence pour chaque classe ; ou elle peut se fonder sur la minimalisation de la distance à des points de référence.

Un dispositif visant à réaliser une reconnaissance automatique doit satisfaire aux impératifs énoncés. De plus, dans bien des cas, la reconnaissance ne pourra s'affranchir d'un certain taux d'erreur, et on a intérêt à mettre en œuvre une stratégie efficace à cet égard. C'est de ce problème que nous allons discuter maintenant.

### *4.1.2. Stratégie et efficacité de la reconnaissance.*

a) *Position du problème.*

Considérons par exemple le cas de la parole. Il y a une forte dispersion dans les paramètres physiques d'un même phonème prononcé par différents locuteurs : un phonème donnera lieu à différents points de $E$ formant un domaine. Les domaines correspondant à des phonèmes différents auront une partie commune. Ainsi pour le cas des voyelles il paraît adéquat de fonder la reconnaissance sur la position des deux premiers formants $F_1$ et $F_2$ : or les domaines du plan $(F_1, F_2)$ correspondant à des voyelles différentes se recoupent [130]. Une catégorisation fondée sur les paramètres $F_1$ et $F_2$ ne se fera donc pas sans erreur : on a intérêt à se rapprocher le plus possible d'un processus de décision « optimal » dans le sens qu'il minimalise les erreurs (ou plus généralement le coût des erreurs, si toutes les erreurs possibles ne sont pas également préjudiciables). Comme on le verra, un tel processus n'est pas toujours facile à mettre en œuvre : aussi comparerons-nous dans des cas particuliers à la catégorisation optimale des catégorisations linéaires (et en particulier celles qui procèdent par maximalisation de la fonction de corrélation croisée), qui sont plus faciles à réaliser artificiellement à l'aide de circuits électriques.

b) *Stratégie optimale.*

On ne peut parler de stratégie optimale qu'au sens d'une « fonction de perte » spécifiant le coût des erreurs. Supposant donnée une telle fonction et supposant connues les propriétés statistiques des signaux et des systèmes de réception, Chow [131] a donné l'expression de la fonction de décision optimale dans un cas assez général ; dans [1], nous avons repris son formalisme en faisant intervenir de plus, dans l'espace $P$ des signaux physiques, les distributions $\sigma_I(S)$ des signaux physiques correspondant à une même classe $I (I = 1, 2, \ldots, K)$ de probabilité *a priori* $p_I$ (cette distribution permettant de traduire, par exemple, la dispersion des paramètres physiques pour un même phonème prononcé par différents locuteurs).

Dans le cas particulier où l'on attribue une perte 1 à toute décision incorrecte et une perte 0 à toute décision correcte, la stratégie optimale consiste à choisir la classe donnant lieu à la plus forte probabilité de production du vecteur $X$ observé dans l'espace $E$.

On peut, pour simplifier, désigner par $\sigma_I\left(\vec{X}\right)$ la distribution dans l'espace $E$ des points correspondant à la classe $I$ (cette distribution pourra décrire à la fois la dispersion des paramètres physiques pour une même classe et les propriétés statistiques du bruit). La règle de décision optimale consiste à choisir la classe $L$ qui donne lieu à :

$$p_L\,\sigma'_L\left(\vec{X}\right) > p_J\,\sigma'_J\left(\vec{X}\right), \qquad \text{pour tout} \qquad J \neq L.$$

Considérons à présent des cas particuliers.

c) *Décision optimale linéaire.*

Supposons d'abord que les distributions $\sigma'_I\left(\vec{X}\right)$ soient gaussiennes et identiques (ne différant que par leur valeur moyenne). Ce cas est important dans

la pratique, car on peut souvent supposer que le bruit introduit une distribution gaussienne. Marill et Green [127] ont montré que dans ce cas la catégorisation optimale est linéaire si l'on suppose tous les $p_I$ égaux : en fait cette hypothèse n'est pas nécessaire pour que la catégorisation optimale soit linéaire [1].

d) *Cas de la corrélation croisée.*

Il est commode de catégoriser en choisissant la classe associée à celle de $K$ configurations de référence qui donne lieu à la plus forte corrélation croisée avec la configuration considérée. On doit pour cela avoir choisi $K$ points de référence $Z_I$ ($I = 1, 2, \ldots, K$) de coordonnées $Z_I^j$ ($j = 1, 2, \ldots, N$). A un point $X$ de coordonnées $X_j$ ($j = 1, 2, \ldots, N$) on associera la classe du point de référence $Z_I$ qui donne lieu à la valeur la plus élevée pour :

$$\Gamma_I = \frac{\sum_{j=1}^{N} Z_I^j X^j}{\sqrt{\sum_{j=1}^{N} (Z_I^j)^2} \sqrt{\sum_{j=1}^{N} (X^j)^2}}.$$

Une catégorisation fondée sur la maximalisation de la fonction de corrélation croisée est équivalente à une catégorisation qui associe à $X$ la catégorie du plus proche de l'un des $K$ points de référence. Si $E$ est un plan, la partition sera alors effectuée par les médiatrices des segments joignant deux à deux les points de référence ; dans le cas général d'un espace $E$ à $N$ dimensions, la partition sera effectuée à l'aide d'hyperplans médiateurs. Inversement il n'est pas possible en général de réaliser une catégorisation linéaire quelconque par corrélation croisée avec des points de référence [1].

La catégorisation par maximalisation de la corrélation croisée n'est donc qu'un cas particulier de catégorisation linéaire. Montrons que dans certains cas on peut réaliser la catégorisation optimale à l'aide de la corrélation : supposons que toutes les probabilités *a priori* $p_I$ soient égales et que toutes les distributions $\sigma_I'\left(\vec{X}\right)$ soient identiques (ne différant que par leurs valeurs moyennes) et non nécessairement gaussiennes. La catégorisation optimale maximalisera $\sigma_I'\left(\vec{X}\right) = \sigma_I'\left(\vec{X} - \vec{M_i} + \vec{M_i}\right)$, soit une fonction $\Sigma_I\left(\vec{X} - \vec{M_i}\right)$. D'après l'hypothèse, cette fonction est indépendante de $I$. Si $\Sigma\left(\vec{Y}\right)$ est une fonction monotone décroissante de $\left|\vec{Y}\right|$, maximaliser $\Sigma\left(\vec{X} - \vec{M_i}\right)$ reviendra à minimaliser $\left|\vec{X} - \vec{M_i}\right|$.

Donc si toutes les catégories sont équiprobables *a priori* et ont mêmes distributions, fonctions décroissantes de $\left|\vec{X} - \vec{M_i}\right|$ (ce qui implique que les distributions soient à symétrie sphérique et que la valeur moyenne $\vec{M_i}$ soit aussi la valeur la plus probable), la catégorisation optimale sera réalisée par minimalisation de distance (on choisit la classe associée à celui des $K$ points de référence qui est le plus proche du point considéré) ou, si l'on préfère, par maximalisation de la corrélation croisée. (Ce résultat suppose une matrice de coût $W_{ij} = 1 - \delta_{ij}$). Les conditions ci-dessus sont réalisées si la distribution est gaussienne avec une matrice de covariance scalaire.

e) *Limites de la corrélation ; filtres de propriétés.*

On conçoit que la reconnaissance des formes par corrélation puisse être efficace si chaque classe est convenablement décrite par une forme-type autour de laquelle les membres de la classe sont assez régulièrement dispersés : le résultat que nous venons d'obtenir montre que l'on peut même atteindre ainsi le processus optimal. Mais on peut avoir affaire à des distributions ayant plusieurs maximums de probabilité et une valeur moyenne dépourvue de signification. Ainsi, il est raisonnable de penser que les distributions des paramètres physiques pour un phonème donné comporteront deux maximums de probabilité correspondant aux voix masculines et féminines [132]. Il peut même être impossible de trouver des formes-types pour certaines classifications : quelle serait la forme-type de la classe des figures qui possèdent la propriété de comprendre un nombre pair de parties disjointes [133] ?

Une reconnaissance de formes peut être effectuée d'après le résultat de mesures sur un phonème physique ; mais on peut supposer un traitement préalable plus élaboré, reconnaissant dans le phonème la présence ou l'absence d'un certain nombre de « propriétés » ou « traits distinctifs » [133].

Une décision ainsi fondée sur une liste de « propriétés » permet d'envisager des reconnaissances de formes plus générales que celles qui procèdent par comparaison à une forme-type.

Considérons un système de reconnaissance possédant $N$ « filtres de propriétés » : une configuration sera représentée par une liste de propriétés, c'est-à-dire par un vecteur $\vec{X}(x^1, x^2, \ldots, x^N)$ où $x^i = 1$ si la configuration jouit de la propriété $P_i$ et $x^i = 0$ dans le cas contraire. En supposant les propriétés indépendantes, nous allons montrer que dans l'hypothèse d'un bruit élevé par rapport au signal (mais n'allant pas jusqu'à « saturer » les filtres), la reconnaissance optimale de configurations connues fait appel à une corrélation, et qu'il en est de même de la classification optimale en catégories définies sur les propriétés de façon probabiliste (c'est-à-dire définies par les probabilités $\sigma_j$ pour qu'un élément de la catégorie $J$ ait la propriété $P_i$).

Dans le cas de configurations connues, nous supposerons que chaque filtre de propriété accomplit sur la configuration $S_J$ une mesure $s_J^i$ et qu'ensuite un dispositif de tout ou rien décide, suivant que la valeur de $s_J^i$ est supérieure ou inférieure à un certain seuil, que la configuration jouit ou non de la propriété $P_i$, ce qui correspond à $x^i$ valent 1 ou 0. Désignons par $\sigma_J^i$ la probabilité pour que $x^i = 1$ en présence du signal $S_J$ : nous supposons ces diverses pro-

babilités indépendantes. Le formalisme sera le même pour les deux cas envisagés.

Le bruit spontané introduit des $s_0^i$ : soit $\sigma_0$ la probabilité pour que le $s_0^i$ correspondant au bruit seul soit au-dessus du seuil (nous supposons cette probabilité constante et indépendante de $i$).

Si l'on désigne par $F(X|S_J)$ la probabilité d'observer le vecteur $X$ quand le signal $S_J$ est présent, et par $F(X|S_0)$ la probabilité d'observer ce vecteur $X$ quand il n'y a que du bruit, la procédure optimale désignera la classe $L$ maximalisant $p_L F(X|S_L)$. Supposons que la probabilité d'avoir $x^i = 1$ à cause du signal soit faible vis-à-vis de celle d'avoir $x^i = 1$ à cause du bruit : $\sigma_L^i - \sigma_0 \ll \sigma_0$. Supposons de plus que la probabilité d'avoir $x_i = 1$ à cause du bruit ne soit pas trop forte et que $\sigma_L^i - \sigma_0 \ll 1 - \sigma_0$. (Cela correspond aux hypothèses d'un faible rapport signal sur bruit et d'un bruit ne « saturant » pas les filtres de propriétés.) Nous avons montré que, sous ces hypothèses, la procédure de décision optimale consiste à maximaliser une fonction où la configuration observée n'intervient que par l'intermédiaire de son produit scalaire (c'est-à-dire de sa corrélation) avec le vecteur $Z_L$. En particulier la procédure de détection optimale d'une seule configuration dans le bruit consiste à calculer le produit $\vec{X}.\vec{Z}$ et à le comparer à un certain seuil. Ces résultats sont applicables à toute reconnaissance de formes effectuée après extraction de propriétés ou d'échantillons binaires indépendants [1].

f) *Méthodes d'apprentissage.*

La catégorisation optimale ne procède qu'exceptionnellement par corrélation : pour la mettre en œuvre dans le cas le plus général, il faut connaître les fonctions $p_I \sigma_I (X)$ (cf. [131] et § 4.1.2*b*). Considérons un cas pratique : supposons que la reconnaissance des formes soit effectuée sur des données auditives ayant préalablement subi une analyse en fréquence par un banc de filtres, comme dans le modèle proposé par Moles (cf. [1] et § 1.2). Ce cas n'a pas seulement un intérêt théorique : une telle analyse est réalisée dans un vocoder à canaux [134] et elle a été utilisée dans diverses expériences de reconnaissance automatique des syllabes [135] ou des chiffres parlés [136, 137]. Soit $N$ le nombre de filtres ; supposons que dans chaque canal l'énergie soit quantifiée à $n$ valeurs. Dans la pratique, $N$ est de 10 à 20, $n$ de 5 à 20. Soit $K$ le nombre de classes : pour spécifier les fonctions $p_I \sigma_I (X)$ il faut $K(n^N - 1)$ nombres [127], soit environ $10^{11}$ nombres pour $K = n = N = 10$. On ne peut donc espérer mettre en œuvre la stratégie optimale indiquée par Chow [131]. Cela montre l'intérêt pratique des techniques d'apprentissage [138, 139, 140, 141], qui utilisent les propriétés statistiques des configurations rencontrées pour améliorer la stratégie de reconnaissance et la faire converger vers des catégorisations linéaires optimales (lesquelles sont, comme nous l'avons vu, au moins aussi bonnes que des classements par corrélation).

## 4.2. Perception auditive et reconnaissance des formes.

### *4.2.1. Analogie entre systèmes sensoriels et systèmes artificiels.*

On peut considérer les systèmes sensoriels comme des systèmes de communication véhiculant de l'information sur l'environnement aux centres cérébraux supérieurs : dans ces centres supérieurs, l'information sensorielle est combinée avec l'information *a priori* pour servir de base à des reconnaissances de forme permettant des décisions sur les événements d'où est issue l'information. Les canaux sensoriels comportent du bruit : ils engendrent une activité spontanée aléatoire et non significative difficile à distinguer de l'information extérieure, et les centres supérieurs opèrent des décisions d'après une information dégradée. Cette hypothèse est corroborée par des mesures de potentiels nerveux [25, 108].

L'intérêt d'un système de communication tient à la réception de messages de certaines classes ; on ne connaît pas exactement *a priori* la structure du message que le système doit transmettre, sinon la communication n'aurait pas de raison d'être. A partir du message reçu on ne peut déduire le message envoyé que si l'on connaît le comportement exact du système : en fait, on peut en prédire le fonctionnement moyen — et traiter les écarts et fluctuations comme du bruit [142]. De là partent les méthodes d'optimalisation pour lesquelles les messages sont considérés comme de nature stochastique : comme nous l'avons vu ci-dessus (cf. §4.1.2*b*), pour un système de reconnaissance à caractéristiques fixes, la stratégie optimale, au sens d'une matrice de coût $W_{ij} = 1 - \delta_{ji}$, est celle qui choisit la catégorie donnant la plus forte probabilité de conduire au vecteur $X$ observé. « L'observateur idéal » doit donc avoir estimé les distributions des paramètres de reconnaissance.

A partir d'une telle conception s'est développé un cadre théorique pour l'étude de certains aspects de la perception auditive : la théorie de la détection [143, 144, 6] interprète les résultats d'expériences de détection auditive des signaux dans le bruit, comparant la performance des sujets humains à celle de « l'observateur idéal ». Mais la théorie de la détection est d'une application limitée. On s'est souvent prévalu de l'analogie entre systèmes sensoriels et systèmes artificiels pour étendre à la perception des résultats valables pour certains systèmes artificiels de reconnaissance [129] et cela peut être fructueux puisque certaines propriétés suggérées par l'analogie ont été vérifiées. Ainsi, les stratégies optimales supposent connues des distributions de probabilité : dans une certaine mesure, l'homme paraît capable d'apprécier ces distributions [145, 146].

De même, l'idée d'une décision fondée sur une liste de propriétés (cf. § 4.2*e*) paraît applicable dans plusieurs cas à la perception auditive. Le cerveau

comprendrait une agrégation d'unités (de « filtres de propriétés »), capables de reconnaître la présence d'une propriété particulière dans l'excitation nerveuse qui l'atteint ; on a des indices anatomiques et physiologiques sur l'existence de telles unités sélectives [147], même à la périphérie du système auditif [148]. La perception paraît décider en fonction d'une liste de propriétés plutôt que par référence à une image idéale, de même qu'un médecin fait son diagnostic en interprétant une liste de « propriétés », un ensemble de signes ou de symptômes présents ou absents, plutôt qu'il ne compare l'état du malade à « l'image type » d'une maladie.

Les « traits distinctifs » de la phonologie pourraient être un exemple de « propriétés » dégagées dans la perception : Jakobson [149] a établi une classification binaire des phonèmes à l'aide d'une dizaine de « traits distinctifs » présents ou absents, ou éventuellement non pertinents (par exemple la nasalité, opposant *m* ou *n* qui possèdent cette propriété à *l* ou *b*). Pour identifier $N$ phonèmes, il suffit de $\log_2 N$ questions binaires indépendantes ; d'après Jakobson, les systèmes phonétiques de toutes les langues pourraient se décrire à l'aide d'une douzaine de traits distinctifs binaires. Or il ne s'agit pas simplement d'un artifice de classement commode sans rapport avec le processus de perception : Miller et Nicely, étudiant les confusions entre seize consonnes anglaises masquées par du bruit blanc, ont trouvé que cinq traits distinctifs étaient perçus plus ou moins indépendamment [150] ; et une expérience récente [151] suggère qu'une consonne est codée dans la mémoire à court terme, non pas comme une unité, mais comme un ensemble de traits distinctifs dont chacun peut être oublié de façon partiellement indépendante.

Enfin l'étude que nous avons effectuée sur les sons de trompette [3, 5] indique que pour identifier l'instrument par son timbre, l'auditeur tient compte de la présence ou l'absence de certaines propriétés.

#### *4.2.2. Originalité de la perception.*

A prendre trop à la lettre des analogies, certes frappantes, entre certains processus de perception et le fonctionnement de certains systèmes artificiels, on risque d'oublier de profondes différences. Un système de reconnaissance de formes est destiné à répertorier automatiquement des signaux suivant des catégories bien définies. Dans la perception, l'affectation des stimuli à des classes spécifiées *a priori* ne joue qu'un rôle partiel ; dans le cas le plus général, l'individu sera confronté à un ensemble de stimuli sans qu'il y ait au départ de classification imposée : il décomposera le tout en parties potentielles, il jaugera les rapports que ces parties ont entre elles, leur organisation dans le tout (cela à plusieurs niveaux imbriqués) pour aboutir à un classement et éventuellement à une reconnaissance. Les « traits distinctifs » éventuels ne sont pas donnés *a priori* : ils sont extraits, dégagés par des opérations perceptives qui dépendent de l'expérience antérieure du sujet [152]. Or cette expérience est très riche : la perception tient compte « en parallèle » d'un nombre considérable de données, et elle opère par une série de tests d'hypothèse, procédant par inférence à partir des indices congruents, et éliminant les indices non congruents — le plus souvent de façon quasiinstantanée et inconsciente. L'élimination spontanée d'indices non congruents est illustrée de façon spectaculaire par l'exemple donné figure 8 ; nous avons montré cette figure quelques

GREVE SUR LE LE TAS

COEUR SUR LA LA MAIN

Fig. 8. — Figure lue incorrectement lors d'une présentation brève.

secondes à vingt sujets en leur demandant de répéter exactement ce qu'ils avaient vu : tous ont répondu « Grève sur le tas, cœur sur la main », omettant la répétition de l'article « le » ou « la » ; nous avons demandé à cinq de ces sujets de s'assurer par un nouveau coup d'œil qu'ils ne s'étaient pas trompés : aucun ne s'est aperçu de l'omission. Nous avons pu donner un exemple comparable (bien que moins frappant) dans le domaine auditif : nous avons demandé à dix sujets d'écouter attentivement, pour tâcher de les comprendre, des signaux de parole de mauvaise qualité, accompagnés d'un léger bruit de fond (suivant les cas, musique, bruit continu, bruit intermittent) : interrogés après l'écoute, deux seulement ont pu préciser la nature du bruit de fond, alors que tous ont pu le faire lorsque leur attention était concentrée vers ce bruit. (Il faut préciser que le signal et le bruit de fond émanaient de deux haut-parleurs distincts.)

Les modèles ont peine à donner une idée des processus de perception et de leur richesse. La perception opère à plusieurs niveaux, elle est apte à passer d'un niveau à un autre, à effectuer, soit une exploration analytique, soit une synthèse de données sensibles [129, 152]. En présence d'une forme distordue, ambiguë, la perception, passant à un autre niveau, pourra tenir compte du « contexte » pour corriger l'erreur ou lever l'ambiguïté : une telle opération fera souvent appel au « sens » du contexte.

Il semble que l'évolution ontogénique de l'individu ait constitué, grâce aux possibilités associatives de la mémoire, une sorte d'image cérébrale, de représentation interne du monde extérieur, sans cesse modifiée et « remise à jour » par la perception ; le sens que nous pouvons donner à un message sensoriel serait lié à cette représentation, en fonction de laquelle les indices sensoriels subiraient un traitement complexe [153, 154].

On peut, dans certains cas très simples, formaliser et réaliser artificiellement des tests « sémantiques » (éliminant par exemple, dans l'interprétation d'un message, des possibilités non cohérentes avec l'état présent de la machine [155]) ; mais, par rapport aux

analyses sémantiques effectuées constamment par le cerveau humain, ce sont des balbutiements ne pouvant avoir qu'une valeur de démonstration : pour un contexte d'une certaine richesse, le nombre de cas possibles est énorme. Bar Hillel [156] considère que le problème de mettre au point un dispositif de traduction entièrement automatique donnant de bons résultats pour un choix de textes étendu est pour l'instant insoluble : il est en effet souvent nécessaire de faire intervenir le sens pour lever l'ambiguïté. Par exemple, « pen » peut signifier « plume » ou « enclos » : le sens impose aux traducteur humain des acceptions différentes dans les deux phrases suivantes :

— « The pen was in the inkstand », la plume était dans l'encrier.

— « The inkstand was in the pen », l'encrier était dans l'enclos.

Le choix du sens correct suppose ici une évaluation spatiale des objets conduisant à une inférence ; dans d'autres cas, il supposera d'autres évaluations plus ou moins complexes, dont l'immense variété traduit la richesse de l'expérience humaine. Aussi ne sait-on pas « coder » le sens de façon utilisable par une machine : cela suppose de nouvelles techniques, peut-être de nouveaux concepts. Les calculateurs actuels, comme l'a fait remarquer Ross Ashby, ne font qu'une chose à la fois, et leurs mémoires sont agencées de façon incommode : dans une mémoire-machine de $10^6$ mots, pour emmagasiner un élément binaire (bit) d'information, il faut dix-neuf éléments binaires pour localiser cette information. On étudie les analogies entre la mémoire humaine et de nouvelles méthodes d'organisation des mémoires artificielles [157] ; des mémoires associatives pourraient aider à lever les difficultés d'adressage et peut-être à envisager une « ontogenèse » de systèmes artificiels de reconnaissance plus prometteuse que les méthodes d'apprentissage statistique [138].

De même qu'un système de traduction automatique, un système de reconnaissance automatique d'écriture ou de parole devrait, pour être d'un usage très large et très sûr, avoir des facultés comparables aux possibilités du cerveau, notamment dans le domaine sémantique. Cela apparaît irréalisable dans l'immédiat : aussi faut-il peut-être demander moins à la machine. C'est dans ce but que McMahon [158] a construit un « sous-ensemble » de la langue anglaise, se lisant comme de l'anglais, mais susceptible d'analyse grammaticale sans ambiguïté (dans l'anglais proprement dit, on est souvent obligé d'avoir recours au sens pour lever les ambiguïtés syntaxiques, comme dans les exemples bien connus suivants : « Time flies like an arrow », « Fruit flies like a peach ». (Le temps fuit comme une flèche, les mouches à fruits aiment une pêche). La traduction automatique de ce « sous-langage » serait beaucoup plus facile que celle du langage lui-même, surtout pour des textes scientifiques au vocabulaire limité et peu ambigu.

Pour illustrer les considérations qui précèdent, nous allons préciser maintenant quelques particularités de la perception humaine dans la compréhension de la parole.

#### *4.2.3. Exemple : remarques sur la compréhension de la parole.*

Signalons que l'idée d'une « analyse par synthèse » dans l'audition, grâce à laquelle nous avons interprété notre expérience de reconnaissance des hauteurs d'un complexe harmonique (cf. § 2.3*c*), intervient également dans une théorie récente de la perception de la parole [159]. Des théories voisines (« proprioceptives ») soutiennent que l'auditeur peut évaluer le signal de parole en estimant l'activité motrice qui l'a produit et en « mimant » cérébralement la production du signal reçu [160].

Nous avons fait remarquer, paragraphe 4.1.2*a*, la diversité des paramètres physiques correspondant à un phonème donné émis par divers locuteurs dans diverses conditions. Or l'auditeur humain est capable d'identifier correctement les phonèmes, dépassant, semble-t-il, les stratégies statistiquement optimales. Pour expliquer cela, Joos a émis l'hypothèse d'une adaptation auditive de l'auditeur à un locuteur particulier, permettant à l'auditeur d'apprécier la relation entre les fréquences des formants alors que ces fréquences peuvent varier d'un locuteur à l'autre pour un même phonème [53]. Cette hypothèse a reçu une confirmation expérimentale [161] : cependant, d'autres expériences indiquent que l'auditeur n'accomplit que de faibles « ajustements » auditifs quand il écoute différents locuteurs [162].

En tout cas, l'hypothèse de Joos [53] est insuffisante, car elle reste sur un plan purement acoustique : or, l'auditeur humain est capable de comprendre correctement un signal de parole distordu et imprécis, car il peut lever l'ambiguïté acoustique par le recours à des informations linguistiques, voire sémantiques. Les résultats des tests d'intelligibilité suivant que les échantillons présentés sont des syllabes dénuées de sens, des mots ou des phrases [163] illustrent bien les facultés humaines à pallier les insuffisances ou les ambiguïtés des indices acoustiques : l'esprit tire parti, quand il le peut, de sa connaissance du vocabulaire de la langue (mot, fréquence de ces mots [146]), et aussi du contexte : probabilités conditionnelles d'observer tel mot après tel autre, syntaxe, sens. Ces trois derniers indices de contexte sont dans une certaine mesure indépendants [164].

Nous avons réalisé une expérience sommaire, mais qui montre que dans deux contextes différents le même signal de parole $X$ peut être compris de deux façons différentes. $X$ est un membre de phrase enregistré — à dessein — avec du bruit de fond (mais sans filtrage). A l'aide de copies (pour la duplication de $X$) et de collages de bande magnétique, nous avons formé deux stimuli $A \equiv Xa$ et $B \equiv Xb$, en adjoignant à $X$ deux membres de phrase différents ($a$ : « de l'église St-Étienne du Mont » ;

*b* : « où l'on pendait les voleurs »). Deux groupes de cinq sujets chacun ont participé à l'expérience, le groupe *A* avec le stimulus *A*, le groupe *B* avec le stimulus *B*. Nous avons demandé aux sujets de noter exactement ce qu'ils entendaient, en leur indiquant que l'expérience avait pour but d'évaluer la gêne due aux défauts d'un système téléphonique. Le stimulus a été répété plusieurs fois. Les cinq sujets du groupe *B* ont tous donné comme réponse : « Avez-vous vu le gibet où l'on pendait les voleurs » ; deux sujets du groupe *A* ont écrit : « Avez-vous vu le gibet de l'église St-Étienne du Mont » et les trois autres : « Avez-vous vu le jubé de l'église St-Étienne du Mont ». On peut sans doute obtenir des résultats plus démonstratifs encore par un soigneux dosage des stimuli (par exemple en procédant à des filtrages passe-bas appropriés, rendant *i* et *u* presque identiques acoustiquement).

On pourra objecter que l'exemple précédent ne fait pas nécessairement intervenir le sens, car il y a probablement un lien statistique entre jubé et église, gibet et pendre, et qu'à partir d'une table de probabilités conditionnelles suffisamment grande, on pourrait mettre au point un système de nature statistique qui lèverait les ambiguïtés d'après les mots du contexte. On peut avoir avantage, dans un système de reconnaissance, à tirer parti de données statistiques de la langue utilisée, comme l'a confirmé une expérience de Denes et Fry [165, 166]. Mais nous emprunterons au domaine de la traduction automatique, où se posent les mêmes problèmes, un exemple montrant les limites d'un système purement statistique. Si l'on suppose que le verbe tirer n'a que deux sens, le sens de traction et le sens d'usage d'une arme à feu, quel sens faut-il lui attribuer dans la phrase suivante : « Le chasseur posa son fusil et tira de son carnier un superbe lièvre qu'il venait de tuer » ? Un dispositif programmé pour décider du sens d'après les mots du contexte, et qui choisirait pour « tirer » le second sens (usage d'une arme à feu dans des contextes de chasse ou de guerre, serait induit en erreur par cette phrase.

On voit que la perception de la parole tient compte d'une multitude d'indices d'origine linguistique ou sémantique aussi bien qu'acoustique (spectre instantanés et leur évolution, rythme, intonation...) ; ces indices interviennent à plusieurs niveaux hiérarchisés, et c'est ce qui permet aux interlocuteurs de se comprendre, malgré les différences physiques entre les signaux qu'ils émettent pour prononcer le même phonème, le même mot ou la même phrase.

Nous dirons à présent quelques mots de la perception de la musique.

#### *4.2.4. Remarques sur la perception de la musique.*

Nous avons indiqué (cf. § 2.4) que dans la perception des hauteurs musicales intervient le conditionnement culturel : et d'abord dans la familiarité qu'il confère avec le milieu sonore, les échelles sonores discontinues (ou gammes) utilisées, qui s'apprennent un peu comme un langage [167, 168]. Le choix des gammes n'est pas arbitraire ; ainsi beaucoup de gammes comportent un nombre d'échelons qui correspond sensiblement à la capacité de classement de la mémoire phosphorescente (cf. § 1.4*d*).

La perception des mélodies apparaît comme une perception de « formes » dans un espace à deux dimensions (hauteurs musicales, temps) symbolisé par la notation musicale occidentale traditionnelle (qui est peut-être le premier système de coordonnées cartésiennes de l'histoire). Les règles de l'écriture musicale occidentale classique définissent des groupes de transformations qui préservent une certaine unité dans les variations d'un motif mélodique : translation (répétition ou transposition), affinité (dilatation ou contraction rythmique ou mélodiques), homothétie, symétrie (renversement mélodique ou récurrence) [169]. Mais la perception de la musique ne se réduit pas à une reconnaissance de formes dans le sens restreint que l'on donne à cette expression dans le domaine technique. L'auditeur ne se borne pas à reconnaître des formes répertoriées à l'avance : la perception de la musique est plutôt une quête de formes latentes au sein du discours musical ; comme nous l'a fait remarquer Benzécri, l'attrait de la perception de la musique tient pour beaucoup à cette recherche même, et le charme est rompu quand on arrive trop bien à cerner et à expliciter les formes.

Comme la perception de la parole, mais d'une manière sans doute plus souple, la perception de la musique est hautement organisée, et enserrée dans une hiérarchie à plusieurs niveaux. La parole utilise une langue pour véhiculer un sens qui est dans une certaine mesure indépendant des signes : rien de tel dans la musique, cependant un style musical particulier possède certaines propriétés proches de celles d'une langue. Tout d'abord, on peut le caractériser par certaines propriétés statistiques [170, 171]. Mais aussi un style musical bien défini paraît posséder une véritable syntaxe, autre sans doute que celle d'une langue, mais qu'il est inadéquat de décrire par un schéma à états finis, comme le montrent les difficultés rencontrées en composition musicale automatique pour organiser une composition globalement et non seulement localement [27]. Le compositeur humain ne se borne pas à filtrer l'aléatoire par des règles de rejet ; comme l'a fait remarquer Babbitt, les règles du contrepoint disent ce qu'il ne faut pas faire plutôt que ce qu'il faut faire [172]. Kassler [173] s'est attaqué à la formalisation de la théorie de Schenker sur la tonalité : selon cette théorie, toute composition tonale peut être engendrée par l'application répétée de « techniques de prolongation » à des « ursatz », brèves séquences de notes postulées tonales constituant une sorte de noyau pour la composition. Kassler cherche à traduire cette théorie dans un langage formel, d'où découlerait un algorithme susceptible de mise en œuvre automatique (par exemple par un calculateur

électronique) : cet algorithme permettrait de décider si une composition donnée est ou non tonale (au sens de Schenker), et aussi de synthétiser des compositions tonales — par des processus non markoviens rappelant ceux indiqués par Chomsky [164] dans le cas de la langue anglaise.

Le conditionnement musical intervient encore au niveau de la syntaxe : ainsi la musique légère qu'on entend un peu partout « conditionne » à une syntaxe tonale une partie de plus en plus importante de l'humanité. On peut rapprocher l'expérience de Meyer [34], dans le domaine de la musique, de celle que nous avons citée, paragraphe 4.2.3, sur le rôle du contexte dans la compréhension de la parole : on pourrait dire, dans l'expérience de Meyer, que les sujets sont influencés par la syntaxe ou le « sens musical » du contexte.

Ces remarques se rapportent à un aspect de la perception musicale, qui suppose des processus mentaux actifs chez l'auditeur. A cette écoute active, organisée, se mêle de façon variable et complexe un aspect plus passif et vague, pour ainsi dire végétatif. Moles a tenté une intéressante distinction entre l'information « sémantique » et l'information « esthétique » [152] : la première serait associée à l'aspect actif et précis de l'écoute musicale, et la seconde à l'aspect passif et vague. Cependant nous sommes ici sur un terrain difficile. La valeur esthétique d'une fugue de Bach écrite sans indications d'instrumentation ne paraît guère dépendre du choix des instruments qui l'exécutent ; et l'on voit mal comment séparer le « sens » de la musique de son pouvoir esthétique : pour l'auditeur, la découverte, la prise de conscience de formes nouvelles paraît un élément important de la jouissance esthétique. Mais nous serions entraînés trop loin à discuter du fondement du sentiment esthétique, que les remarques précédentes effleurent à peine [174].

## CONCLUSION

Partant d'un modèle fonctionnel de l'audition, nous avons étudié certains aspects de la perception des sons. Nous avons mis l'accent sur l'intérêt des modèles de l'audition, mais aussi sur la complexité et la pluralité des mécanismes auditifs humains. Nous avons examiné comment on pouvait concilier les diverses théories de la perception des hauteurs sonores ; nos expériences confirment la dépendance des mécanismes de perception des hauteurs vis à vis du contexte ou du conditionnement antérieur du sujet : ces particularités ressortent également d'un examen de la perception des hauteurs musicales. Dans l'audition peuvent intervenir des opérations de corrélation : leur étude nous a conduit à proposer une méthode d'analyse en fréquence applicable sur calculateurs. Les processus de la perception auditive sont très élaborés et relèvent du problème de la reconnaissance des formes : nous avons abordé ce problème du point de vue de sa résolution dans un système technique, puis nous avons fait ressortir l'originalité des processus de reconnaissance des formes chez l'homme : ces processus font appel à des indices très divers qui traduisent la richesse de l'expérience humaine et qui ne sont guère accessibles actuellement à une machine, même programmée avec raffinement.

Au cours de cet exposé, nous avons parfois souligné les limitations des calculateurs et les insuffisances des modèles : il ne faut prendre cela que comme une mise en garde contre le danger de prendre trop à la lettre certaines analogies apparentes. Il reste que c'est en bonne partie grâce aux modèles qu'on a pu mieux circonscrire l'originalité des processus humains de perception, qui réside dans cette prodigieuse richesse de l'expérience humaine dont les modèles ne peuvent rendre compte. Pris avec précaution, les modèles aident à préciser l'idée qu'on peut se faire des mécanismes élaborés de l'audition ; grâce à leurs facultés de simulation et leur capacité de traitement, les calculateurs les rendent plus féconds.

D'autre part la complexité de la perception auditive n'est pas rebelle à toute analyse scientifique. Les modèles suggèrent des expériences, pour lesquelles les calculateurs peuvent constituer de précieux outils. Mentionnons simplement que de puissantes techniques relevant de l'analyse factorielle, et commodes à mettre en œuvre grâce aux calculateurs, permettent de dégager, à partir de jugements subjectifs, des propriétés significatives, et d'isoler des dimensions psychologiques sous-jacentes [175, 176, 177].

Nous tenons à exprimer notre reconnaissance envers le Professeur Grivet, pour son accueil et son soutien constant ; le Professeur Moles, pour ses avis précieux ; le maître André Jolivet, pour son vigoureux encouragement ; divers membres des Bell Telephone Laboratories et plusieurs musiciens, particulièrement Michel Philippot, James Tenney et les regrettés Suzanne Demarquez et Edgard Varèse, pour de stimulantes discussions ; l'Ingénieur général Libois et l'Ingénieur des télécommunications Ferrieu, qui nous ont permis de terminer ce travail lors d'un fructueux séjour au Centre National d'Études des Télécommunications ; le Pr Benzécri, qui a suggéré maintes améliorations du manuscrit.

*Manuscrit reçu le* 11 *mai* 1967.

*Manuscrit révisé reçu le* 11 *juillet* 1967.

## BIBLIOGRAPHIE

[1] Risset (J. C.). Sur l'analyse, la synthèse et la perception des sons, étudiées à l'aide de calculateurs électroniques. *Thèse*, Faculté des Sciences d'Orsay (1967). Manuscrit intégral déposé à la bibliothèque du CNET.

[2] Moles (A.). Aspect temporel de l'intégration du message linguistique et modèles de l'audition. In A. Moles et B. Vallancien, Communications et langages, *Gauthier-Villars*, Paris (1963), pp. 81-89.

[3] Risset (J. C.). Computer study of trumpet tones. (Étude des sons de trompette à l'aide de calculateur.), (1966), *J. Acoust. Soc. Am.* (1965), **38**, p. 912.

[4] Anonyme. Computer synthesizes trumpet sounds. (Synthèse des sons de trompette au moyen d'un calculateur.) *Bell Laboratories Record*, (1965), **43**, p. 417.

[5] Risset (J. C.). *a*) Spectre des sons de trompette. *C. R. Ac. Sci. B Fr.* (1966), **262**, pp. 1245-1248 ; *b*) Analyse des sons de trompette à l'aide d'un calculateur électronique, *C. R. Ac. Sci. B Fr.* (1966), **262**, pp. 1650-1653 ; *c*) Synthèse des sons à l'aide de calculateurs électroniques appliquée à l'étude des sons de trompette, *C. R. Ac. Sci. B. Fr.* (1966), **263**, pp. 111-114.

[6] Mathews (M. V.). Signal detection for human auditory perception. (Détection du signal dans la perception auditive chez l'homme.) *J. Wiley*, New-York (1963), pp. 349-361.

[7] Bekesy (G. V.). Experiments in hearing. (Audition expérimentale.) *McGraw Hill*, New York (1960).

[8] Flanagan (J. L.). Models for approximating basilar membrane displacement. (Modèles approchés du déplacement de la membrane basilaire.) *Bell Syst. Techn. J.* U. S. A. (1960), **39**, pp. 1163-1191 ; *Bell Syst. Techn. J.* U. S. A. (1962), **41**, pp. 959-1009.

[9] Flanagan (J. L.). Computer simulation of basilar membrane displacement. (Simulation sur calculateur du déplacement de la membrane basilaire.) Paper H 26 in *Proc. of the 4th Internat. Congress on Acoustics*, Copenhagen (1962).

[10] Flanagan (J. L.), Guttman (N.). On the pitch of periodic pulses. (Hauteur d'un son formé d'impulsions périodiques.) *J. Acoust. Soc. Am.* (1960), **32**, pp. 1308-1319.

[11] Flanagan (J.-L.), Guttman (N.). Pitch of periodic pulses without fundamental component. (Hauteur d'un son formé d'impulsions périodiques sans composante fondamentale.) *J. Acoust. Soc. Am.* (1960), **32**, pp. 1319-1328.

[12] Guttman (N.), Flanagan (J.-L.). Pitch of high pass filtered pulse trains. (Hauteur du son formé de trains d'impulsions ayant traversé un filtre passe-haut.) *J. Acoust. Soc. Am.* (1964), **36**, pp. 757-765.

[13] Katsuki (Y.). Neural mechanisms of auditory sensation in cats. (Mécanismes nerveux de la perception auditive chez le chat.) Sensory Communications, *J. Wiley*, New York (1961), pp. 561-583.

[14] Fessard (A.). The role of neuronal networks in sensory communications within the brain. (Rôle des réseaux de neurones pour la transmission d'informations sensorielles à l'intérieur du cerveau.) Sensory Communication. *J. Wiley*, New York (1961), pp. 585-607.

[15] Furman (G. G.), Frishkopf (L. S.). Model of neural inhibition in the mammalian cochlea. (Modèle de l'inhibition neurale dons la cochlée des mammifères). *J. Acoust. Soc. Am.* (1964), **36**, pp. 2194-2201.

[16] Khanna (S. M.), Noback (C. R.). A model for processing in the nervous system. (Un modèle pour les processus du système nerveux. *Proc. New York Acad. of Sciences* (1963), **91**, pp. 516-521.

[17] Licklider (J. C. R.). On psychophysiological models. (Modèles psychophysiologiques.) Sensory Communications. *J. Wiley*, New-York (1961).

[18] Helmholtz (H. F.). On the sensations of tone. (Sur les sensations auditives.) New York (1954).

[19] Stevens (S. S.), Davis (H.). Hearing (Audition.) *J. Wiley*, New York (1952).

[20] Wever (E. G.). Theory of hearing. (Théorie de l'audition.) *J. Wiley et Sons*, New York (1952).

[21] Boudreau (J. C.). Neural volleying : upper frequency limits detectable in the auditory system. (Théorie de la « volée » des fibres nerveuses : détection dans le système auditif de limites supérieures en fréquence). *Nature* G. B. (1965), **208**, p. 1237.

[22] Licklider (J. C. R.). A duplex theory of pitch perception. (Un modèle « duplex » de la perception de la hauteur.) *Experientia*, Suisse (1951), **7**, pp. 128-133.

[23] Licklider (J. C. R.). Auditory frequency analysis. **In** C. Cherry., *Proc. 3d London Symposium on Inform. Butterworth*, London (1956), pp. 253-268.

[24] Miller (G. A.), Licklider (J. C. R.). The intelligibility of interrupted speech. (L'intelligibilité de la parole interrompue.) *J. Acoust. Soc. Am.* (1950) **22**, pp. 167-173.

[25] Eijkman (E.), Vendrik (A. J. H.). Detection theory applied to the absolute sensitivity of the sensory system. (Théorie de la détection appliquée à la sensibilité absolue du système sensoriel.) *Biophys. J.*, U. S. A. (1963), **3**, pp. 65-78.

[26] Moles (A.). Étude et représentation de la note complexe en acoustique musicale. *Ann. Télécommunic.* Fr. (1952), **7**, n° 11, pp. 430-438.

[27] Risset (J.-C.). Musique, électronique et théorie de l'information. *L'Onde Electrique* Fr. (1964), **451** pp. 1055-1063.

[28] Plomp (R. M.), Levelt (W. J. M.). Tonal consonance and critical bandwidth. (Consonance tonique et bande passante critique.) *J. Acoust. Soc. Am.* (1965), **38**, pp. 548-560.

[29] Frances (R.). La perception de la musique. *Vrin*, Paris (1958).

[30] Elfner (L.). Systematic shifts in the judgment of octaves of high frequencies. (Erreurs systématiques dans l'appréciation des octaves de hautes fréquences.) *J. Acoust. Soc. Am.* (1964), **36**, pp. 270-276.

[31] Dubois (T.). Traité d'harmonie théorique et pratique. *Heugel*, Paris (1901).

[32] Pikler (A. G.). Information analysis of power-spectrum model of hearing. (Analyse de l'information contenue dans un modèle de spectre de puissance de l'audition.) *J. Acoust. Soc. Am.* (1955), **27**, p. 208.

[33] Boomsliter (P. C.), Creel (W.). Ratio relationships in melody. (Relations rationnelles dans la mélodie.) *J. Acoust. Soc. Am.* (1962), **34**, p. 1276.

[34] Meyer (M. F.). Listeners can be seducted to perceive the paradoxical ratio 51/87 as either one or another truly melodical interval. (Des auditeurs croient percevoir le rapport paradoxal 51/87 comme l'un ou l'autre de deux intervalles vraiment mélodiques.) *J. Acoust. Soc. Am.* (1962), **34**, pp. 1277.

[35] Lafon (J. C.). La reconnaissance phonétique et sa mesure. *Ann. Télécommunic.* Fr. (1960), **15**, pp. 27-37.

[36] Shannon (C. E.). Communication in the presence of noise. (Transmission en présence de bruit.) *Proc. I. R. E.* U. S. A. (1949), **37**, pp. 10-21.

[37] Moore (E. F.), Shannon (C. E.). Reliable circuits using less reliable relays. (Circuits fiables utilisant des relais de moindre qualité.) *J. Franklin Inst.* U, S, A. (1956), **26**, pp. 191-208 et pp. 281-297.

[38] Van Bergeijk (W. A.). Analogy of the external spiral innervation of the cochlea. (Analogie de l'innervation spiralée extérieure de la cochlée.) *Kybernetik* Dstch. (1961), **3**, pp. 14-19.

[39] Harmon (L. P.) et Wolfe (R. M.). An electronic model of a nerve cell. (Un modèle électronique d'une cellule nerveuse.) *Semiconductor product* U. S. A. (1959), **2**, p. 36.

[40] Stevens (S. S.). Is there a quantal threshold ? (Existe-t-il un seuil quantique ?) Sensory Communication. *J. Wiley*, New York (1961), pp. 806-813.

[41] Stevens (S. S.), Morgan (C. T.), Volkmann (J.). Theory of the neural quantum in the discrimination of loudness and pitch. (Théorie du quantum neural dans la discrimination de l'intensité et de la hauteur du son.) *Americ. J. Psychol.* (1941), **54**, pp. 315-335.

[42] Barlow (H. B.). Comment on neural quanta. (Remarques sur le « quantum neural ».) Sensory communication. *J. Wiley*, New York (1961), pp. 786-790.

[43] Corso (J. F.). The neural quantum theory of sensory discrimination. (La théorie du « quantum neural » dans la discrimination sensorielle.) *Psychol. Bull.* U. S. A. (1956), **53**, pp. 371-393.

[44] Guilford (J. P.). Psychometric methods. (Méthodes phsychométriques.) *McGraw Hill*, New York (1954).

[45] Norman (D. A.). Stochastic learning and a quantal model of signal detection. (Étude stochastique en modèle quantique de la détection du signal.) **In** *Joint Autom. Control Conf. Amer. Inst. Electr. Engrs*, New York (1962), pp. 35-40.

[46] Benzecri [J. P.]. La psychophysique : histoire et critique de la notion de seuil. *Publications de l'Institut de Statistique de l'Université de Paris* (1966), p. 153.

[47] Jacobson (H.). Information and the human ear. (L'information et l'oreille humaine.) *J. Acoust. Soc. Am.* (1951), **23**, pp. 463-471.

[48] Pollack (I.). The information of elementary auditory displays. (Le contenu d'information des formes auditives élémentaires.) *J. Acoust. Soc. Am.* (1952), **24**, pp. 745-749.

[49] Pimonow (L.). La parole synthétique et son application dans la correction auditive. *Ann. Télécommunic.* Fr. (1965), **20**, nos 7-8, pp. 151-171.

[50] Dixon Ward (W.). Information and absolute pitch. (Information et hauteur absolue d'un son.) *J. Acoust. Soc. Am.* (1953), **25**, p. 833.

[51] Harris (G. G.). Periodicity perception by using gated noise. (Perception d'une périodicité dans un bruit modulé en créneaux.) *J. Acoust. Soc. Am.* (1963), **35**, pp. 1229-1233.

[52] Miller (G. A.). The perception of short bursts of noise. (La perception de bruits intenses et courts.) *J. Acoust. Soc. Am.* (1948), **20**, pp. 160-170.

[53] Joos (M.). Acoustic phonetics. (Phonétique acoustique.) *Supplement to Language* U. S. A. (1948), **24**.

[54] Gabor (D.). Theory of communication. (Théorie de la transmission.) *J. Inst. Elec. Engrs*, G. B. (1946), **93**, pp. 429-457.

[55] Zwislocki (J.), Theory of temporal auditory summation. (Théorie de l'additurité auditive temporelle.) *J. Acoust. Soc. Am.* (1960), **32**, pp. 1040-1060.

[56] Hirsch (I.). Auditory perception of temporal order. (Perception auditive d'ordre temporel.) *J. Acoust. Soc. Am.* (1959), **31**, p. 759.

[57] Pikler (A. G.). Musical transfer functions and processed music. (Fonctions de transfert musicales.) *I. R. E. Trans.* A. A. U. S. A. (1962), **10**, no 2, pp. 47-52.

[58] Tenney (J. C.). Discriminability of differences in the rise-time of a tone. (Discrimination de différences dans les temps de montée d'un son pur.) *J. Acoust. Soc. Am.* (1962), **34**, p. 739.

[59] Pollack (I.). Onset discrimination for white noise. (Repérage de l'instant d'apparition d'un bruit blanc.) *J. Acoust. Soc. Am.* (1962), **34**, p. 1973.

[60] Miller (G. A.). The magical number seven plus or minus two : some limits on our capacity for processing information. (Le nombre magique sept plus ou moins deux : quelques limites à notre capacité de traitement de l'information.) *Psychol. Rev.* U. S. A. (1956), **63**, pp. 81-97.

[61] Miller (G. A.). A note on the remarkable memory of man. (Note sur la mémoire remarquable de l'homme.) *I. R. E. Trans.* E. C. (1957), **6**, no 3, pp. 194-195.

[62] Crossman (E. R.). Information and serial order in the human immediate memory. (Rangement des informations sur la mémoire à court terme de l'homme.) In C. Cherry, Proc. 4th London Symposium on Inf. Th., *Butterworth*, London (1960).

[63] Sidley (N. A.), Winograd (E.), Bedarf (E. W.). Stimulus-identification overlap in learning to identify complex sounds. (Confusion entre stimuli différents dans l'apprentissage d'un son complexe.) *J. Acoust. Soc. Am.* (1965), **38**, pp. 11-13.

[64] Sperling (G.). The information available in brief visual presentations. (Contenu d'information de présentations visuelles rapides.) *Psychol. Monogr.* U. S. A. (1960), **74**, no 498.

[65] Shepard (R. N.). Recognition memory for words, sentences, and pictures. (Mémoires de reconnaissance des mots, des phrases et des images.) *Bell Telephone Lab. Techn.* Memorandum MM 65-1221-4 U. S. A. (1965).

[66] Moller (A. A.). Acoustic reflex in man. (Réflexes acoustiques de l'homme.) *J. Acoust. Soc. Am.* (1962), **34**, pp. 1524-1534.

[67] Tasaki (I.), Davis (H.), Legouix (J. P.). Space-time pattern of the cochlear microphonics (guinea pig) as recorded by differential electrodes. (Programmes espace-temps de potentiel microphonique cochléaire enregistré par des électrodes différentielles.) *J. Acoust. Soc. Am.* (1952), **24**, pp. 502-519.

[68] Rutherford (W.). A new theory of hearing. (Une nouvelle théorie de l'audition.) *J. Anat. Physiol.* (1886), **21**, p. 166.

[69] Goldstein (M. H.). Jr., Kiang (N. Y.-S.). Synchrony of neural activity in electrical responses evoked by transient acoustic stimuli. (Synchronisme des activités nerveuses dans les réponses électriques à des stimuli acoustiques transitoires.) *J. Acoust. Soc. Am.* (1958), **30**, pp. 107-114.

[70] Kiang (N. Y.-S.), Goldstein (M. H.) Jr., Peake (W. T.). Temporal coding of neural responses to acoustic stimuli. (Codage temporel des réponses nerveuses à des stimuli acoustiques.) *I. R. E. Trans.*, I. T. (1962), **8**, pp. 113-119.

[71] Schuknecht (H. F.), Sutton (S. S.). Hearing losses after experimental lesions in basal coil of cochlea. (Baisse de l'audition après lésion expérimentale de la membrane basilaire de la cochlée.) *Arch. Otolaryng.* U. S. A. (1953), **57**, p. 129.

[72] HUGGINS (W. H.). Quarterly Progress Report, Res. Lab Electron., *Mass. Inst. Technol.* (octobre 1950), pp. 54-59.

[73] ALLANSON (J. T.), NEWELL (A. F.). Subjective responses to tones modulated simultaneously in both amplitude and frequency. (Réponses subjectives à des sons simultanément modulés en amplitude et en fréquence.) *J. Sound Vib.* G. B. (1966), **3**, pp. 135-146.

[74] TONNDORF (J.). Time/frequency analysis along the partition of cochlear models : a modified place concept. (Analyse harmonique et temporelle le long de la partition des modèles de la cochlée ; modifications de la théorie de la localisation.) *J. Acoust. Soc. Am.* (1962), **34**, pp. 1337-1350.

[75] BOUASSE (H.). Bases physiques de la musique. *Gauthier-Villars*, Paris (1906).

[76] CARRÉ (R.), LANCIA (R.), PAILLE (J.), GSELL (R.). Étude et réalisation d'un détecteur de mélodie pour analyse de la parole. *L'Onde électrique*, Fr (1963), **434**, pp. 556-562.

[77] SCHOUTEN (J.-F.), RITSMA (R. J.), LOPES CARDOZO (B.). Pitch of the residue. (Hauteurs des fondamentaux résiduels.) *J. Acoust. Soc. Am.* (1962), **34**, pp. 1418-1424.

[78] SMALL (A. M.), CAMPBELL (R. A.). Masking of pulsed tones by bands of noise. (Impulsions sonores masquées par des bandes de bruit.) *J. Acoust. Soc. Am.* (1961), **33**, pp. 1570-1576.

[79] PLOMP (R.). The ear as a frequency analyzer. (L'oreille en tant qu'analyseur de fréquence.) *J. Acoust. Soc. Am.* (1964), **36**, pp. 1628-1636.

[80] FLETCHER (H.). Auditory patterns. (Diagrammes auditifs.) *Rev. Modern. Phys.* U. S. A. (1940), **12**, pp. 47-65.

[81] ZWICKER (E.), FLOTTORP (G.), STEVENS (S. S.). Critical bandwidth in loudness summation. (Bande passante critique dans l'additivité des intensités.) *J. Acoust. Soc. Am.* (1957), **29**, pp. 548-557.

[82] GREENWOOD (D. D.). Critical bandwith and the frequency coordinates of the basilar membrane. (La bande passante critique et les coordonnées de fréquence de la membrane basilaire.) *J. Acoust. Soc. Am.* (1961), **33**, pp. 1344-1356.

[83] RITSMA (R. J.). Existence region of the tonal residue. (Région d'existence des résidus toniques). *J. Acoust. Soc. Am.* (1962), **34**, pp. 1224-1229.

[84] DIXON WARD (W.). Diplacusis and auditory theories. (Théorie de l'audition) *J. Acoust. Soc. Am.* (1963), **35**, pp. 1746-1747.

[85] BEKESY (G. V.). Similarities between hearing and skin sensations. (Similitudes entre les sensations auditives et tactiles.) *Psychol. Rev.* (1959), **66**, pp. 1-22.

[86] DAVIS (H.). Peripheral coding of auditory information. (Codage périphérique des informations auditives.) Sensory Communications *J. Wiley* (1961) New-York pp. 319-141.

[87] PIMONOW (L.). Contribution à l'étude du problème de la sélectivité tonale de l'oreille. *Acustica* Dtsch. (1959), **9**, pp. 345-352.

[88] KARPLUS (H. B.). Correlation hypothesis to explain the fine frequency discrimination of the ear. (Hypothèse de corrélation pour expliquer la discrimination fine en fréquence de l'oreille.) *J. Acoust. Soc. Am.* (1963), **35**, p. 809.

[89] STEVENS (S. S.). The relation of pitch to intensity. (La relation entre la hauteur et l'intensité.) *J. Acoust. Soc. Am.* (1935), **6**, p. 150.

[90] MORGAN (C. T.), GARNER (W. R.), GALAMBOS (R.). Pitch and intensity. (Hauteur et intensité.) *J. Acoust. Soc. Am.* (1951), **23**, pp. 658-664.

[91] COHEN (A.). Further investigations on the effects of intensity upon the pitch of pure tones. (Contribution à l'étude des effets de l'intensité sur la hauteur des sons purs.) *J. Acoust. Soc. Am.* (1961), **33**, pp. 1363-1373.

[92] CRAMER (E. M.), HUGGINS (W. H.). Creation of pitch through binaural interaction. (Création de la hauteur d'un son par interaction entre les deux oreilles.) *J. Acoust. Soc. Am.* (1958), **30** pp. 413-417.

[93] KENDALL (B. R. F.). Resolving-power multiplier for mass spectrometer. (Amélioration du pouvoir de résolution d'un spectromètre de masse.) U. S. A. *Rev. Sci. Instr.* (1962), **33**, pp. 30-37.

[94] ESCUDIÉ (B.), MORGUE (M.), CHERY (R.). Simulateur électrique analysant des mesures complexes par augmentation du pouvoir de résolution. *C. R. Acad. Sci.*, Fr. (1964), **258**, pp. 852-855.

[95] BISHOP (G. R.), GROSSETETE (B.), RISSET (J. C.). Étude de l'électrodésintégration et de la photodésintégration de l'Oxygène 16. *J. Phys. Rad.* Fr. (1962), **23**, pp. 31-42.

[96] MATHEWS (M. V.). An acoustic compiler for music and psychological stimuli. (Un compilateur acoustique pour la musique et les stimuli psychologiques.) *Bell Syst. Techn. J.* U. S. A. (1961), **40**, pp. 677-694.

[97] WRIGHT (H. N.). Audibility of switching transients. (Audibilités des transitoires en commutation.) *J. Acoust. Soc. Am.* (1960), **32**, p. 138.

[98] FLANAGAN (J. L.), GUTTMAN (N.), WATSON (B.). Pitch of periodic pulses with non-uniform amplitudes. (Hauteur d'un son formé d'impulsions périodiques d'amplitudes non uniformes.) *J. Acoust. Soc. Am.* (1962), **34**, p. 738.

[99] GUTTMAN (N.), PRUZANSKY (S.). Lower limits of pitch and musical pitch. (Limites inférieures de la hauteur et de la hauteur en musique.) *J. Speech et Hearing Res.* U. S. A. (1962), **5**, pp. 207-214.

[100] SMITH (S.). Diphlophonie und Luft-Schall- Explosionen. (La diphonie et les consonnes explosives.) *Arch. Ohren-, Nasen- u. Kehlkopfheilk. ver. Z ; Hals-, Nasen- u. Ohrenheilk* Dtsch. (1958), **173**, pp. 504-508.

[101] LIEBERMANN (P.). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. (Mesures acoustiques de la périodicité fondamentale de larynx normaux et pathologiques.) *J. Acoust. Soc. Am.* (1963), **35**, pp. 344-355.

[102] RISSET (J. C.). Modèle numérique de la reconnaissance des formes par corrélation et application à la fonction auditive. *C. R. Acad. Sci.*, Fr. (1964), **259**, pp. 1219-1221.

[103] POLLACK (I.). Ohm's acoustical law and short-term auditory memory. (Lois d'Ohm et mémoire d'audition à court terme.) *J. Acoust. Soc. Am.* (1964), **36**, pp. 2340-2345.

[104] REVESZ (G.). Zur Grundlegung der Tonpsychologie. (Base de la psychologie des sons.) *Viet* Leipzig (1913).

[105] SHEPARD (R. N.). Circularity in judgments of relative pitch. (Sur le rangement des hauteurs relatives selon une échelle circulaire.) *J. Acoust. Soc. Am.* (1964), **36**, pp. 2346-2353.

[106] SMALL (A. M.). An objective analysis of artistic violin performance. (Analyse objective des performances artistiques du violon.) *Univ. Iowa Studies in the psychology of music* U. S. A. (1936), **4**, pp. 172-231.

[107] CHIARUCCI (H.), REIBEL (G.). Rapport entre la hauteur et le fondamental d'un son musical.

*Audiologie intern.* Néderl. (1966), **5**, nº 3, pp. 400-407.

[108] Brazier (M. A. B.), Killam (K. F.), Hance (A. J.). The reactivity of the nervous system in the light of the past history of the organism (Réactivité du système nerveux en fonction de l'histoire antérieure de l'organisme), Sensory communications *J. Wiley*, New York (1961), pp. 699-716.

[109] Dixon Ward (W.). Subjective musical pitch. (Hauteur musicale subjective.) *J. Acoust. Soc. Am.* (1954), **26**, pp. 369-380.

[110] Bachem (A.). Time factors in relative and absolute pitch determination. (Facteurs temporels dans les déterminations des hauteurs relatives et absolues.) *J. Acoust. Soc. Am.* (1954), **26**, pp. 751-753.

[111] Tanner (W. P.) Jr., Rivette (C. L.). Experimental study of tone deafness. (Étude expérimentale de la surdité.) *J. Acoust. Soc. Am.* (1964), **36**, pp. 1465-1467.

[112] Lee (Y. W.), Cheatham (T. P.), Wiesner (J. B.). Application of correlation analysis to the detection of periodic signals in noise. (Application des méthodes de corrélation à la détection de signaux périodiques dans du bruit.) *Proc. I. R. E.*, U. S. A. (1950), **38**, pp. 1165-1171.

[113] Fano (R. M.). Short-time autocorrelation functions and power spectra. (Fonction d'autocorrélation à court terme et spectre de puissance.). *J. Acoust. Soc. Am.* (1950), **22**, pp. 546-550.

[114] Schroeder (M. R.), Atal (B. S.). Generalized short-time power spectra and autocorrelation functions. (Spectres de puissance et fonctions d'autocorrélation à court terme). *J. Acoust.* pp. 1679-1683. *Soc. Am.* (1962), **34**,

[115] Blackman (R. M.), Tukey (J. W.). The measurement of power spectra. (Mesure des spectres de puissance). *Dover*, New York, (1959).

[116] Atal (B. S.), Schroeder (M. R.), Kuttruff (K. H.). Perception of coloration in filtered gaussian noise-short-time spectral analysis by the ear. (Perception par l'oreille de la coloration d'un bruit gaussien filtré après une analyse spectrale à court terme). *Proc. of the* 4th *I. C A. Congress*, H 31. Copenhagen (août 1962).

[117] Wente (E. C.). Characteristics of sound transmissions in rooms. (Caractéristiques de la transmission des sons dans les salles.) *J. Acoust. Soc. Am.* (1935), **7**, p. 123.

[118] Van der Pol (R.). Frequency modulation. (Modulation de fréquence.) *Proc. I. R. E.*, U. S. A. (1930), **18**, p. 1194.

[119] Mathes (R. C.), Miller (R. L.). Phase effects in monaural perception. (Effets de phase dans la perception par une seule oreille.) *J. Acoust. Soc. Am.* (1947), **19**, pp. 780-797.

[120] Pierce (J. R.). Some work on hearing. (Travaux sur l'audition.) *Amer. Scientist* (1960), **48**, pp. 40-45.

[121] Craig (J. H.), Jeffress (L. A.). Effect of phase on the quality of a 2-component tone. (Effets de phase sur la qualité d'un son à deux composantes.) *J. Acoust. Soc. Am.* (1962), **34**, pp. 1752-1760.

[122] Hebb (D. O.). Organization of behavior. (Organisation du comportement.) *J. Wiley*, New York (1959).

[123] Levy (J. C.). Un modèle de réseau figurant le système nerveux. *Onde Électrique* Fr. (1963), **433**, 435-451.

[124] Elias (P.). A note on the misuse of " digital " in neurophysiology. (Du mauvais usage du mot « digital » en neurophysiologie.) Sensory Communication. *J. Wiley*, New York (1961), pp. 794-795.

[125] Gaito (J.). D. N. A. and R. N. A. as memory molecules. (L'A. D. N. et l'A. R. N. considérées comme supports moléculaires de la mémoire.) *Psychol. Rev.* U. S. A. (1963), **70**, p. 471.

[126] Martin (T. B.), Talavage (J. J.), Putzrath (F. L.). Electronic model of the human auditory system. (Modèle électronique du système auditif de l'homme.) *Proc. of the* 4th *I. C. A. Congress*, Copenhagen, (août 1962), G 25.

[127] Marill (T.), Green (D. M.). Statistical recognition functions and the design of pattern recognizers. (Fonctions de reconnaissance statistique et mise au point de processus de reconnaissance.) *I. R. E. Trans.*, E. C. U. S. A. (1960), **9**, nº 4 ,pp. 472-477.

[128] Grivet (P.), Blaquière (A.). Le bruit de fond, *Masson*, Paris (1958), 426 p.

[129] McCulloch (W. S.), Pitts (W. H.). How we know universals. The perception of auditory and visual forms. (Comment nous connaissons l'univers. Perception des formes audibles et visibles.) *Bull. Math. Biophysics* U. S. A. (1947), **9**, pp. 127-147.

[130] Peterson (G. H.), Barney (H. L.). Control methods used in a study of the vowels. (Méthodes de contrôle utilisées dans l'étude des voyelles.) *J. Acoust. Soc. Am.* (1952), **24**, pp. 175-184.

[131] Chow (C. K.). An optimum character recognition system using decision functions. (Un système de reconnaissance à caractère optimal utilisant les fonctions de décision.) *I. R. E. Trans.*, E. C. U. S. A. (1957), **6**, nº 4, pp. 247-254.

[132] Denes (P.), Mathews (M. V.). Spoken digit recognition using time-frequency pattern matching. (Reconnaissance de chiffres parlés par comparaison de diagrammes temps-fréquence.) *J. Acoust. Soc. Am.* (1960), **32**, pp. 1450-1455.

[133] Minsky (M.). Steps toward artificial intelligence. (Approches de l'intelligence artificielle.) *Proc. I. R. E.*, U. S. A. (1961), **49**, pp. 8-30.

[134] Flanagan (J. L.). Speech analysis, synthesis and perception. (Analyse, synthèse et perception de la parole.) *Springer-Verlag*, Berlin (1965), 247 p.

[135] Olson (H. F.), Belar (H.). Syllabe analyzer, coder and synthesizer for the transmission of speech. (Analyseur, codeur et synthétiseur de syllabes pour la transmission de la parole.) *Proc. I. R. E. Trans., A. U.* (1962), **10**, pp. 11-17.

[136] Davis (K. H.), Biddulph (R.), Balashek (S.). Automatic recognition of spoken digits. (Reconnaissance automatique de chiffres parlés) *J. Acoust. Soc. Am.* (1952), **24**, pp. 637-642.

[137] Sholtz (P. N.), Bakis (R.). Spoken digit recognition using vowel-consonant segmentation. (Reconnaissance des chiffres parlés par segmentation entre les voyelles et les consonnes.) *J. Acoust. Soc. Am.* (1962), **34**, pp. 1-5.

[138] Nilsson (N. J.). Learning machines. (Machines capables d'apprentissage.) *McGraw Hill*, New York (1965).

[139] Rosenblatt (F.). Principles of neurodynamics : perceptrons and the theory of brain mechanism. (Principes de dynamique nerveuse : les « perceptrons » et la théorie des mécanismes cérébraux.) *Spartan Books*, Washington (1961).

[140] Castan (S.), Perennou (J.). Séparation linéaire de formes par apprentissage. *C. R. Acad. Sci.* Fr. (1966), **262**, pp. 52-54.

[141] Scuder (H. J.), Probability of error of some adaptive pattern recognition machines. (Proba-

bilité d'erreurs de certaines machines à reconnaître des formes et capables d'apprentissage). *I. E. E. E. Trans.* I.T. (1965), **11**, pp. 363-371.

[142] PIERCE (J. R.). Proposal for an explanation of limens of loudness. (Tentative d'explication des seuils d'audition.) *J. Acoust. Soc. Am.* (1958), **30**, pp. 418-420.

[143] TANNER (W. P.), SWETS (J. A.). A decision-making theory of human detection. (Théorie de la décision et perception chez l'homme.) *Psychol. Rev.* U. S. A. (1954), **61**, pp. 401-409.

[144] GREEN (D. M.). Psychoacoustics and detection theory. (Psycho-acoustique et théorie de la détection.) *J. Acoust. Soc. Am.* (1960), **32**, pp. 1189-1203.

[145] UTTLEY (A. M.). The conditional probability of signals in the nervous system. (La probabilité conditionnelle de signaux dans le système nerveux.) Radar Research Establishment Memorandum nº 1109 (1955). *Rev. Mod. Phys.* U. S. A. (1959), **31**, p, 547.

[146] SAVIN (H. B.). Word frequency effect and errors in the perception of speech. (Effets de la fréquence des mots et erreurs dans la perception de la parole.) *J. Acoust. Soc. Am.* (1963), **35**, pp. 202-206.

[147] HUBEL (D.). The visual cortex of the brain. (Le cortex visuel du cerveau.) *Scientif. Amer.* (1963), **209**, nº 11, pp. 54-47. Cf. aussi Sensory Communications *J. Wiley* (1961) New-York pp. 403-436 et pp. 757-776.

[148] FRISHKOPF (L.), GOLDSTEIN (M. H. Jr.), Responses to acoustic stimuli from single units in the eighthnerve of the bull-frog. (Réponses d'unités du huitième nerf de la grenouille à des stimuli acoustiques.) *J. Acoust. Soc. Am.* (1963), **35**, pp. 1219-1228.

[149] JAKOBSON (R.), FANT (C. G. M.), HALLE (M.). Preliminaries to speech analysis. (Introduction à l'analyse de la parole.) Techn. Report nº 13, *Acoustics Lab. Mass. Inst. Technol.*, Cambridge, Mass. (1952).

[150] MILLER (G. A.), NICELY (P. E.). An analysis of perceptual confusions among some English consonants. (Analyse de la confusion de perception de quelques consonnes anglaises.) *J. Acoust. Soc. Am.* (1955), **27**, pp. 338-352.

[151] WICKELGREN (W. A.). Distinctive features and errors in short-term memory for English consonants. (Différences et erreurs dans la mémoire à court terme des consonnes anglaises.) *J. Acoust. Soc. Am.* (1966), **39**, pp. 388-398.

[152] MOLES (A.). Théorie de l'information et perception esthétique. *Flammarion*, Paris (1958).

[153] YOUNG (J. Z.). A model of the brain. (Un modèle du cerveau.) *Clarendon Press*, Oxford (1965).

[154] MCKAY (D. M.). Ways of looking at perception. (Manières de considérer la perception.) *Proc. of Symp. on models for the perception of speech and visual form.* Boston, Mass. (1964).

[155] ROSENSTEIN (M.). A theory of audition. (Une théorie de l'audition.) *I. E. E. E. Internation. Conv. Record*, part. 2 U. S. A. (1966), pp. 157-165.

[156] BAR-HILLEL (Y.). The present status of automatic translation of languages. (État actuel de la traduction automatique des langues.) In F. ALT, Advances in Computers, vol. 1, *Academic Press*, New York, U. S. A. (1960), pp. 91-163.

[157] MATHEWS (M. V.). Scatter memory model for human memory. (De la mémoire humaine considérée comme mémoire aléatoire.) *Bell Teleph. Lab. Techn. Memor.* MM 64-122-1 U. S. A. (1964).

[158] MCMAHON (L. E.). A fundamentally analyzable simplified English : FASE. (Une langue anglaise simplifiée et complètement analysable : le FASE.) *Bell. Teleph. Lab. Techn. Memor.* MM 65-1221-7 U. S. A. (1965).

[159] STEVENS (K. N.). Toward a model for speech recognition. (Recherche d'un modèle de reconnaissance de la parole.) *J. Acoust. Soc. Am.* (1960), **32**, pp. 47-55.

[160] LIBERMAN (A. M.), COOPER (F. S.), HARRIS (K. S.), MCNEILLAGE (P. F.). A motor theory of speech perception. (Une théorie motrice de la perception du langage.) *Proc. of the Speech Communication Seminar*, Stockholm 29, (août 1, Sept. 1962).

[161] LADEFOGED (P.), BROADBENT (D. E.). Information conveyed by vowels. (Informations transmises par les voyelles.) *J. Acoust. Soc. Am.* (1957), **29**, pp. 98-104.

[162] CREELMAN (C. D.). Case of the unknown talker. (Problème du locuteur inconnu.) *J. Acoust. Soc. Am.* (1957), **29**, pp. 655-656.

[163] KRYTER (K. D.). Methods for the calculation and use of the articulation index. (Méthode de calcul et utilisation du taux d'articulation.) *J. Acoust. Soc. Am.* (1962), **34**, pp. 1689-1697.

[164] CHOMSKY (N.). Syntaxic structures. (Structures de la syntaxe.) *S-Gravenhage-Mouton et Co*, Amsterdam (1957).

[165] FRY (D. B.). Perception and recognition in speech. (Perception et reconnaissance de la parole.) *Roman Jakobson* (M. HALLE et *al.*, ed), pp. 169-173, *Mouton*, La Hague (1956).

[166] DENES (P.). The design and operation of the mechanical speech recognizer at University College, London. (La conception et le fonctionnement de la machine à reconnaître la parole de l'University college à Londres.) *J. Brit. I. R. E.* (1959), **10**, pp. 219-229.

[167] VAN ESBROECK (J.), MONTFORT (F.). Qu'est-ce que jouer juste ? *Lumière*, Bruxelles, (1946).

[168] COSTERE (E.). Mort ou transfiguration de l'harmonie. *Presses Universitaires de France*, Paris (1962).

[169] SCHÖNBERG (A.). Harmonielehre. (Leçons d'harmonie.) *Universal*, Wien (1911).

[170] FUCKS (W.). Mathematical analysis of formal structure in music. (Analyse mathématique de la structure formelle en musique.) *Proc. I. R. E. Trans.* I. T. U. S. A. (1962), 8, nº 5, pp. 225-228.

[171] SHANNON (C. E.). Prediction and entropy of printed English. (L'entropie de l'anglais écrit.) *Bell. Syst. Techn. J.* U. S. A. (1951), **30**, pp. 50-64.

[172] PIERCE (J. R.), MATHEWS (M. V.), RISSET (J. C.). Further experiments on the use of the computer in connection with music. (Contributions à l'utilisation de calculateurs pour l'étude de la musique.) *Gravesaner Blätter* 27/28 Nederl. (1965), pp. 85-97.

[173] KASSLER (M.). A report of work, directed towards explication of Schenker's theory of tonality. (Rapport sur un travail essayant d'expliquer la théorie de la tonalité de Shenker.) *Princeton University*, Princeton, *N. J.* U. S. A. (1962).

[174] SCHLOEZER (B. DE). Introduction à J.-S. BACH, *Ed. N. R. F.*, Paris (1947).

[175] BENZECRI (J. P.). Analyse factorielle. *Public. de l'Institut de Statist. de l'Univ. de Paris* (1964), **13**, 235. p.

[176] OSGOOD (C. E.), SUGY (G. C.), TANNENBAUM (P. H.). The measurement of meaning. (Mesure de la signification). Illinois Press, Urbana, Ill. (1957), *Psychol. Bull.* U. S. A. (1957), **53**, pp. 470-538.

[177] SOLOMON (L. N.). Semantic approach to the perception of complex sounds. (Approche sémantique de la perception des sons complexes.) *J. Acoust. Soc. Am.* (1958), **30**, pp. 421-425.

Imprimerie TAFFIN-LEFORT, à Lille (France). — *Published in France*